

2018 IEEE 14th International Conference on e-Science (e-Science 2018)

**Amsterdam, Netherlands
29 October – 1 November 2018**



**IEEE Catalog Number: CFP1806A-POD
ISBN: 978-1-5386-9157-1**

**Copyright © 2018 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:	CFP1806A-POD
ISBN (Print-On-Demand):	978-1-5386-9157-1
ISBN (Online):	978-1-5386-9156-4

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

2018 IEEE 14th International Conference on e-Science e-Science 2018

Table of Contents

Message from the Program Chairs	xxi
Organizing Committee	xxiii
International Steering and Supervisory Committee	xxvi
Program Committees	xxvii

International Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE 6.1, 2018)

Talks I

The Role of Data Stewardship in Software Sustainability and Reproducibility	1
<i>Maria J. Cruz (Delft University of Technology), Shalini Kurapati (Delft University of Technology), and Yasemin Turkyilmaz-van der Velden (Delft University of Technology)</i>	
Talk to Me: A Case Study on Coordinating Expertise in Large-Scale Scientific Software Projects	9
<i>Reed Milewicz (Sandia National Laboratories) and Elaine Raybourn (Sandia National Laboratories)</i>	
Lesson Development for Open Source Software Best Practices Adoption	19
<i>Mateusz Kuzak (Dutch Techcentre for Life Sciences), Jen Harrow (ELIXIR Hub Wellcome Genome Campus, Hinxton, UK), Rafael C. Jimenez (ELIXIR Hub Wellcome Genome Campus, Hinxton, UK), Paula Andrea Martinez (ELIXIR Belgium), Fotis E. Psomopoulos (Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki), Radka Svobodová Vaeková (Central European Institute of Technology (CEITEC), Brno), and Allegra Via (ELIXIR Italy, National Research Council of Italy (CNR), Institute of Molecular Biology and Pathology (IBPM), Italy)</i>	

Talks II

Reflections from a Decade of Running CCPForge	21
<i>Catherine Jones (Science and Technology Facilities Council), Alan Kyffin (Science and Technology Facilities Council), and Gemma Poulter (Science and Technology Facilities Council)</i>	

Painting the Picture of Software Impact with the Research Software Directory	23
<i>Jurriaan H. Spaaks (Netherlands eScience Center), Tom Klaver (Netherlands eScience Center), Stefan Verhoeven (Netherlands eScience Center), Jason Maassen (Netherlands eScience Center), Tom Bakker (Netherlands eScience Center), Atze van der Ploeg (Netherlands eScience Center), Ben van Werkhoven (Netherlands eScience Center), Willem van Hage (Netherlands eScience Center), and Rob V. van Nieuwpoort (Netherlands eScience Center)</i>	

Talks III

Mapping the Research Software Sustainability Space	25
<i>Stephan Druskat (Humboldt-Universität zu Berlin) and Daniel S. Katz (University of Illinois Urbana-Champaign)</i>	
Building a Sustainable Structure for Research Software Engineering Activities	31
<i>Jeremy Cohen (Imperial College London), Daniel S. Katz (University of Illinois Urbana-Champaign), Michelle Barker (Australian Research Data Commons), Robert Haines (University of Manchester), and Neil Chue Hong (Software Sustainability Institute, University of Edinburgh)</i>	
Research Software Discovery: An Overview	33
<i>Alexander Struck (Cluster of Excellence Image Knowledge Gestaltung An Interdisciplinary Laboratory at Humboldt Universität zu Berlin)</i>	
Survey on Research Software Engineering in the Netherlands	38
<i>Ben van Werkhoven (Netherlands eScience Center), Tom Bakker (Netherlands eScience Center), Olivier Philippe (Software Sustainability Institute), and Simon Hettrick (Software Sustainability Institute)</i>	

Workshop on Research Objects (RO2018)

The Archive and Package (arcp) URI Scheme	40
<i>Stian Soiland-Reyes (The University of Manchester) and Marcos Cáceres (Mozilla Corporation)</i>	
Preserving Reproducibility: Provenance and Executable Containers in DataONE Data Packages	45
<i>Bryce Mecum (University of California, Santa Barbara), Matthew B. Jones (University of California, Santa Barbara), Dave Vieglaiss (University of Kansas), and Craig Willis (University of Illinois)</i>	
A Research Object-Based Toolkit to Support the Earth Science Research Lifecycle	50
<i>Raul Palma (Poznan Supercomputing and Networking Center), Andres Garcia-Silva (Expert System), Jose Manuel Gomez-Perez (Expert System), and Marcin Krystek (Poznan Supercomputing and Networking Center)</i>	

Tuesday 30 October: Plenary Session

Digital Methods in Holocaust Studies: The European Holocaust Research Infrastructure	58
<i>Daan de Leeuw (NIOD Institute for War, Holocaust and Genocide Studies), Mike Bryant (Department of Digital Humanities, King's College London), Michal Frankl (Masaryk Institute and Archives of the Czech Academy of Sciences), Ivelina Nikolova (Ontotext), and Vladimir Alexiev (Ontotext)</i>	
Catching Toad Calls in the Cloud: Commodity Edge Computing for Flexible Analysis of Big Sound Data	67
<i>Paul Roe (Dr), Meriem Ferroudj (Queensland University of Technology), Michael Towsey (Queensland University of Technology), and Lin Schwarzkopf (James Cook University)</i>	
FATBIRD: A Tool for Flight and Trajectories Analyses of Birds	75
<i>Daniyal Kazempour (Ludwig-Maximilians-Universität), Anna Beer (Ludwig-Maximilians-Universität), Friederike Herzog (Bavarian Society for the Protection of Birds (LBV), Hilpoltstein, Germany), Daniel Kaltenthaler (Ludwig-Maximilians-Universität), Johannes-Y. Lohrer (Ludwig-Maximilians-Universität), and Thomas Seidl (Ludwig-Maximilians-Universität)</i>	
Nanopublications: A Growing Resource of Provenance-Centric Scientific Linked Data	83
<i>Tobias Kuhn (Vrije Universiteit Amsterdam), Albert Meroño-Peñuela (Vrije Universiteit Amsterdam), Alexander Malic (Maastricht University), Jorrit H. Poelen (Oakland), Allen H. Hurlbert (University of North Carolina), Emilio Centeno Ortiz (Universitat Pompeu Fabra), Laura I. Furlong (Universitat Pompeu Fabra), Núria Queralt-Rosinach (The Scripps Research Institute), Christine Chichester (Datafair.xyz), Juan M. Banda (Georgia State University), Egon Willighagen (Maastricht University), Friederike Ehrhart (Maastricht University), Chris Evelo (Maastricht University), Tareq B. Malas (Leiden University Medical Center), and Michel Dumontier (Maastricht University)</i>	
ScienceSearch: Enabling Search through Automatic Metadata Generation	93
<i>Gonzalo P. Rodrigo (Lawrence Berkeley National Laboratory), Matt Henderson (Lawrence Berkeley National Laboratory), Gunther H. Weber (Lawrence Berkeley National Laboratory), Colin Ophus (Lawrence Berkeley National Laboratory), Katie Antypas (Lawrence Berkeley National Laboratory), and Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)</i>	
Curation of Image Data for Medical Research	105
<i>Lasse Wollatz (University of Southampton), Mark Scott (University of Southampton), Steven J. Johnston (University of Southampton), Peter M. Lackie (University of Southampton), and Simon J. Cox (University of Southampton)</i>	
How FAIR Can you Get? Image Retrieval as a Use Case to Calculate FAIR Metrics	114
<i>Tobias Weber (Leibniz Supercomputing Centre) and Dieter Kranzlmüller (Leibniz Supercomputing Centre)</i>	
Specimens as Research Objects: Reconciliation Across Distributed Repositories to Enable Metadata Propagation	125
<i>Nicky Nicolson (Royal Botanic Gardens, Kew and Brunel University), Alan Paton (Royal Botanic Gardens, Kew), Sarah Phillips (Royal Botanic Gardens, Kew), and Allan Tucker (Brunel University, London)</i>	

Wednesday 31 October: Plenary Session

Fast and Reproducible LOFAR Workflows with AGLOW	136
<i>Alexandar Mechev (Leiden Observatory), Raymond Oonk (Leiden Observatory), Timothy Shimwell (ASTRON), Aske Plaat (Leiden Institute of Advanced Computer Science), Huib Intema (Leiden Observatory), and Huub Rottgerin (Leiden Observatory)</i>	
Visual Programming Languages for Programmers with Dyslexia: An Experiment	145
<i>José Luis Fuertes Castro (Universidad Politécnica de Madrid), Luis Fernando González Alvarán (Universidad Politécnica de Madrid - Politécnico Colombiano Jaime Isaza Cadavid), and Loïc Antonio Martínez Normand (Universidad Politécnica de Madrid)</i>	
An Algebra for Robust Workflow Transformations	156
<i>Nicholas Hazekamp (University of Notre Dame) and Douglas Thain (University of Notre Dame)</i>	
Orchestral: A Lightweight Framework for Parallel Simulations of Cell-Cell Communication	168
<i>Adrien Coulier (Uppsala University) and Andreas Hellander (Uppsala University)</i>	
Pilot-Streaming: A Stream Processing Framework for High-Performance Computing	177
<i>Georgios Chantzialexiou (Rutgers University), Andre Luckow (Clemson University, Ludwig Maximilian University), and Shantenu Jha (Rutgers University, Brookhaven National Laboratory)</i>	
Concurrent and Adaptive Extreme Scale Binding Free Energy Calculations	189
<i>Jumana Dakka (Rutgers University), Kristof Farkas-Pall (University College London), Matteo Turilli (Rutgers University), David W. Wright (University College London), Peter V. Coveney (University College London), and Shantenu Jha (Rutgers University)</i>	
Educational Outreach & Stakeholder Role Evolution in a Cyberinfrastructure Project	201
<i>David P. Randall (University of Washington), Drew Paine (Lawrence Berkeley National Laboratory), and Charlotte P. Lee (University of Washington)</i>	
A Survey of Software Metric Use in Research Software Development	212
<i>Nasir U. Eisty (University of Alabama), George K. Thiruvathukal (Loyola University Chicago), and Jeffrey C. Carver (University of Alabama)</i>	
Building NDStore Through Hierarchical Storage Management and Microservice Processing	223
<i>Kunal Lillaney (Johns Hopkins University), Dean Kleissas (Johns Hopkins University Applied Physics Laboratory), Alexander Eusman (Johns Hopkins University), Eric Perlman (Johns Hopkins University), William Gray Roncal (Johns Hopkins University Applied Physics Laboratory), Joshua T. Vogelstein (Johns Hopkins University), and Randal Burns (Johns Hopkins University)</i>	
Designing Scientific SPARQL Queries Using Autocompletion by Snippets	234
<i>Karima Rafe (BorderCloud), Serge Abiteboul (Ecole Normale Supérieure, Inria), Sarah Cohen-Boulakia (LRI, U. Paris-Sud, U. Paris-Saclay), and Bastien Rance (APHP G. Pompidou, INSERM CRC, U. Paris Descartes)</i>	

Big Provenance Stream Processing for Data Intensive Computations	245
<i>Isuru Suriarachchi (Indiana University), Sachith Withana (Indiana University), and Beth Plale (Indiana University)</i>	
Skluma: An Extensible Metadata Extraction Pipeline for Disorganized Data	256
<i>Tyler J. Skluzacek (University of Chicago), Rohan Kumar (University of Chicago), Ryan Chard (Argonne National Laboratory), Galen Harrison (University of Chicago), Paul Beckman (University of Chicago), Kyle Chard (University of Chicago; Argonne National Laboratory), and Ian Foster (University of Chicago; Argonne National Laboratory)</i>	

Focused Sessions

Weather & Climate Science in the Digital Era

Message from the eScience 2018 Program Committee Chairs for the Focused Session on Weather & Climate Science in the Digital Era	267
<i>Peter Bauer (European Centre for Medium-Range Weather Forecasts, United Kingdom) and Martine de Vos (Netherlands eScience Center, The Netherlands)</i>	
A Hybrid-Resolution Earth System Model	268
<i>Marc Stringer (University of Reading), Colin Jones (University of Leeds), Richard Hill (Hadley Centre for Climate Science), Mohit Dalvi (Hadley Centre for Climate Science), Colin Johnson (Hadley Centre for Climate Science), and Jeremy Walton (Hadley Centre for Climate Science)</i>	
Resolving Clouds in a Global Atmosphere Model - A Multiscale Approach with Nested Models	270
<i>Fredrik Jansson (Centrum Wiskunde & Informatica), Gijs van den Oord (Netherlands eScience Center), Pier Siebesma (KNMI and Delft University of Technology), and Daan Crommelin (Centrum Wiskunde & Informatica and University of Amsterdam)</i>	
Increasing Parallelism in Climate Models Via Additional Component Concurrency	271
<i>Jörg Behrens (Deutsches Klimarechenzentrum), Joachim Biercamp (Deutsches Klimarechenzentrum), Hendryk Bockelmann (Deutsches Klimarechenzentrum), and Philipp Neumann (Deutsches Klimarechenzentrum)</i>	
Extracting Flood Maps from Social Media for Assimilation	272
<i>Etienne Brangbour (LIST), Pierrick Bruneau (LIST), and Stéphane Marchand-Maillet (University of Geneva)</i>	
Toward a Cloud Ecosystem for Modeling as a Service	274
<i>Mohan Ramamurthy (University Corporation for Atmospheric Research)</i>	
Machine Learning for Applied Weather Prediction	276
<i>Sue Ellen Haupt (National Center for Atmospheric Research), Jim Cowie (National Center for Atmospheric Research), Seth Linden (National Center for Atmospheric Research), Tyler McCandless (National Center for Atmospheric Research), Branko Kosovic (National Center for Atmospheric Research), and Stefano Alessandrini (National Center for Atmospheric Research)</i>	

Visibility Prediction Based on Kilometric NWP Model Outputs Using Machine-Learning Regression	278
<i>Driss Bari (CNRMSI/SMN Direction de la Météorologie Nationale)</i>	
Weather Reanalysis on an Urban Scale using WRF	279
<i>Ronald van Haren (Netherlands eScience Center), Sytse Koopmans (Wageningen University and Research), Gert-Jan Steeneveld (Wageningen University and Research), Natalie Theeuwes (University of Reading), Remko Uijlenhoet (Wageningen University and Research), and Albert A.M. Holtslag (Wageningen University and Research)</i>	
Detecting Probability of Ice Formation on Overhead Lines of the Dutch Railway Network	281
<i>Irene Garcia-Marti (Royal Netherlands Meteorological Institute (KNMI)), Gerard van der Schrier (Royal Netherlands Meteorological Institute (KNMI)), Jan Willem Noteboom (Royal Netherlands Meteorological Institute (KNMI)), and Paul Diks (ProRail)</i>	
A Web Service Architecture for Objective Station Classification Purposes	283
<i>Martin G. Schultz (Jülich Supercomputing Centre), Sander Apweiler (Jülich Supercomputing Centre), Jan Vogelsang (Jülich Supercomputing Centre), Björn Hagemeier (Jülich Supercomputing Centre), Felix Kleinert (Jülich Supercomputing Centre), and Daniel Mallmann (Jülich Supercomputing Centre)</i>	

Data Handling and Analytics for Health

Message from the eScience 2018 Program Committee Chairs for the Focused Session on Data Handling and Analytics for Health	285
<i>Jaap Heringa (VU University Amsterdam) and Vincent van Hees (Netherlands eScience Center)</i>	
Gross Motor Activity Patterns in Depression and Anxiety	286
<i>Sonia Difrancesco (Amsterdam Public Health Research Institute), Vadim Zipunnikov (Johns Hopkins University), Robert A. Schoevers (University of Groningen), Harriëtte Riese (University of Groningen), Niki Antypa (Leiden University Medical Center), Brenda W.J.H. Penninx (Amsterdam Public Health Research Institute), Kathleen R. Merikangas (National Institute of Mental Health), Albert A. M. van Hemert (Leiden University Medical Center), and Femke Lamers (Amsterdam Public Health Research Institute)</i>	
NordScreen - An Interactive Tool for Presenting Cervical Cancer Screening Indicators in the Nordic Countries	287
<i>Veli-Matti Partanen (Finnish Cancer Registry), Maiju Pankakoski (Finnish Cancer Registry), Zurab Bzhalava (Karolinska Institutet), Piret Veerus (National Institute for Health), Ahti Anttila (Finnish Cancer Registry), Tytti Sarkeala (Finnish Cancer Registry), Ameli Tropé (Cancer Registry of Norway), Stefan Lönnberg (Finnish Cancer Registry), Sirpa Heinävaara (Finnish Cancer Registry), Joakim Dillner (Karolinska Institutet), and Ágúst Ingi Ágústsson (Icelandic Cancer Society)</i>	

Stratifying Cervical Cancer Risk with Registry Data	288
<i>Nicholas Baltzer (Uppsala University), Mari Nygård (The Cancer Registry of Norway), Karin Sundström (Karolinska Institute), Joakim Dillner (Karolinska Institute), Jan Nygård (The Cancer Registry of Norway), and Jan Komorowski (Uppsala University)</i>	
Automating the Placement of Time Series Models for IoT Healthcare Applications	290
<i>Lauren Roberts (Newcastle University), Peter Michalák (Newcastle University), Sarah Heaps (Newcastle University), Michael Trenell (Newcastle University), Darren Wilkinson (Newcastle University), and Paul Watson (Newcastle University)</i>	
Semantically Enriched Literature Search Combining Text Mining, QSPR and Ontologies in Scientific Workflows	292
<i>Magnus Palmblad (Leiden University Medical Center)</i>	
Differences in the Commonly used Genotype Imputation Algorithms and Their Imputation Accuracy Estimates	293
<i>Kalle Pärn (University of Helsinki), Matti Pirinen (University of Helsinki), Mart Kals (University of Tartu), Reedik Mägi (University of Tartu), Veikko Salomaa (National Institute for Health and Welfare), Michael Boehnke (University of Michigan), Ira Hall (Washington University School of Medicine), Nathan Stitzel (Washington University School of Medicine), Nelson Freimer (University of California Los Angeles), Mark Daly (Broad Institute of MIT and Harvard, Massachusetts General Hospital and Harvard Medical School), Aarno Palotie (University of Helsinki, Broad Institute of MIT and Harvard), Samuli Ripatti (University of Helsinki), and Priit Palta (University of Tartu)</i>	
Mining Events Preceding a Cancer Diagnosis	295
<i>Rebecka Weegar (Stockholm University)</i>	
The Impact of Social Versus Individual Learning for Agents' Risk Perception During Epidemics	297
<i>Shaheen A. Abdulkareem (University of Twente), Ellen-Wien Augustijn (University of Twente), Katarzyna Musial (University of Technology), Yaseen T. Mustafa (University of Zakho (UoZ)), and Tatiana Filatova (University of Technology)</i>	
Workflows Orchestrating Workflows: Thousands of Queries and Their Fault Tolerance Using APIs of Omics Web Resources	299
<i>Yassene Mohammed (Leiden University Medical Center)</i>	
Machine Learning for Multi-Omics Data Integration and Variant Pathogenicity Estimation	301
<i>Shuang Li (University of Groningen and University Medical Center Groningen), K. Joeri van der Velde (University of Groningen and University Medical Center Groningen), and Morris A. Swertz (University of Groningen and University Medical Center Groningen)</i>	
Remote Cloud-Based Automated Stroke Rehabilitation Assessment Using Wearables	302
<i>Shane Halloran (School of Mathematics, Statistics and Physics), Jian Qing Shi (School of Mathematics, Statistics and Physics), Yu Guan (Open Lab), Xi Chen (School of Mathematics, Statistics and Physics), Michael Dunne-Willows (School of Mathematics, Statistics and Physics), and Janet Eyre (Newcastle University)</i>	

A Portable and Scalable Workflow for Detecting Structural Variants in Whole-Genome Sequencing Data	303
<i>Arnold Kuzniar (Netherlands eScience Center), Jason Maassen (Netherlands eScience Center), Stefan Verhoeven (Netherlands eScience Center), Luca Santuari (University Medical Center), Carl Shneider (University Medical Center), Wigard Kloosterman (University Medical Center), and Jeroen de Ridder (University Medical Center)</i>	
Analytics Pipeline for Left Ventricle Segmentation and Volume Estimation on Cardiac MRI Using Deep Learning	305
<i>Mai H. Nguyen (University of California, San Diego), Ehab Abdelmaguid (University of California, San Diego), Jolene Huang (University of California, San Diego), Sanjay Kenchareddy (University of California, San Diego), Disha Singla (University of California, San Diego), Laura Wilke (University of California, San Diego), Marcus Bobar (University of California, San Diego), Eric D. Carruth (University of California, San Diego), Dylan Uys (University of California, San Diego), Ilkay Altintas (University of California, San Diego), Evan D. Muse (Scripps Translational Science Institute), Giorgio Quer (Scripps Translational Science Institute), and Steven Steinhubl (Scripps Translational Science Institute)</i>	

Advances in eScience for the Humanities and Social Sciences

Message from the eScience 2018 Program Committee Chairs for the Focused Session on Advances in eScience for the Humanities and Social Sciences	307
<i>Claes de Vreese (University of Amsterdam) and Carlos Martinez-Ortiz (Netherlands eScience Center)</i>	
Filter and Annotate: Towards Automatic Identification of Genuine Metaphoricity	308
<i>Erik-Lân Do Dinh (Ubiquitous Knowledge Processing (UKP) Lab Technische Universität Darmstadt), Iryna Gurevych (Ubiquitous Knowledge Processing (UKP) Lab Technische Universität Darmstadt), and Petra Gehring (Institute of Philosophy Technische Universität Darmstadt)</i>	
Extracting Theory from Black Boxes: Using Machine Vision APIs in Communication Research	310
<i>Theo Araujo (Amsterdam School of Communication Research (ASCoR) University of Amsterdam), Irina Lock (Amsterdam School of Communication Research (ASCoR) University of Amsterdam), and Bob Van De Velde (Informatics Institute University of Amsterdam)</i>	
Understanding Evolving Communities in Transnational Board Interlock Networks	312
<i>Dafne E. Van Kuppevelt (Netherlands eScience Center), Frank W. Takes (Universiteit van Amsterdam), and Eelke M. Heemskerk (Universiteit van Amsterdam)</i>	
TI-One: Active Research Data Management in a Modern Philosophy Department	314
<i>Gioele Barabucci (University of Cologne), Mark Eschweiler (University of Cologne), and Andreas Speer (University of Cologne)</i>	
Software Engineering and Digital Research Infrastructures	316
<i>Carsten Thiel (CESSDA ERIC), Dieter Van Uytvanck (CLARIN ERIC), and Tibor Kálmán (GWDG/DARIAH)</i>	

Automatically Detecting Incivility in Online Discussions of News Media	318
<i>Johannes Daxenberger (Technische Universität Darmstadt), Marc Ziegele (Heinrich-Heine-Universität Düsseldorf), Iryna Gurevych (Technische Universität Darmstadt), and Oliver Quiring (Johannes Gutenberg-Universität Mainz)</i>	
Social Network-Epistemology	320
<i>M. Alfano (TU Delft), S. Cunningham (TU Delft), W. Meulemans (TU Eindhoven), I. Rutter (University of Passau), M. Sondag (TU Eindhoven), B. Speckmann (TU Eindhoven), and E. Sullivan (TU Delft)</i>	
Democratizing Ancient Mesopotamian Research through Digital Scholarship	322
<i>Raquel Alegre (University College London), Anastasis Georgoulas (University College London), Stuart Grieve (Queen Mary University of London), and Eleanor Robson (University College London)</i>	
Using Facial Expressions of Students for Detecting Levels of Intrinsic Motivation	323
<i>Pedro Bispo Santos (Technische Universität Darmstadt), Caroline Verena Wahle (Universität Koblenz-Landau), and Iryna Gurevych (Technische Universität Darmstadt)</i>	
Bringing Data Science to Qualitative Analysis	325
<i>You-Wei Cheah (Lawrence Berkeley National Laboratory), Drew Paine (Lawrence Berkeley National Laboratory), Devarshi Ghoshal (Lawrence Berkeley National Laboratory), and Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory)</i>	
Linking Text and Knowledge Using the INCEpTION Annotation Platform	327
<i>Richard Eckart De Castilho (Technische Universität Darmstadt), Jan-Christoph Klie (Technische Universität Darmstadt), Naveen Kumar (Technische Universität Darmstadt), Beto Boullosa (Technische Universität Darmstadt), and Iryna Gurevych (Technische Universität Darmstadt)</i>	
INCA: Infrastructure for Content Analysis	329
<i>Damian Trilling (University of Amsterdam), Bob Van De Velde (University of Amsterdam), Anne C. Kroon (University of Amsterdam), Felicia Löcherbach (University of Amsterdam), Theo Araujo (University of Amsterdam), Joanna Strycharz (University of Amsterdam), Tamara Raats (University of Amsterdam), Lisa De Klerk (University of Amsterdam), and Jeroen G.F. Jonkman (University of Amsterdam)</i>	

Exascale Computing for High-Energy Physics

Message from the eScience 2018 Program Committee Chairs for the Focused Session on Exascale Computing for High-Energy Physics	331
<i>Jeff Templon (Nikhef, The Netherlands) and Yifat Dzigan (Netherlands eScience Center, The Netherlands)</i>	

Understanding the Performance of a Prototype of a WLCG Data Lake for HL-LHC	332
<i>Jaroslava Schovancová (CERN, the European Organization for Nuclear Research), Simone Campana (CERN, the European Organization for Nuclear Research), Xavier Espinal Curull (CERN, the European Organization for Nuclear Research), Maria Girone (CERN, the European Organization for Nuclear Research), Ivan Kadochnikov (Joint Institute for Nuclear Research), and Gavin John McCance (CERN, the European Organization for Nuclear Research)</i>	
Modelling High-Energy Physics Data Transfers	334
<i>Joaquin Bogado (Universidad Nacional de La Plata), Fernando Monticelli (Universidad Nacional de La Plata), Javier Diaz (Universidad Nacional de La Plata), Mario Lassnig (CERN), and Ilija Vukotic (University of Chicago)</i>	
Distributed and On-demand Cache for CMS Experiment at LHC	336
<i>Diego Ciangottini (INFN), Daniele Spiga (INFN), Tommaso Boccali (INFN), Giacinto Donvito (INFN), Daniele Cesini (INFN), Giuseppe Bagliesi (INFN), Enrico Mazzone (INFN), and Antonio Falabella (INFN)</i>	
Fine-Grained Processing Towards HL-LHC Computing in ATLAS	338
<i>Doug Benjamin (Duke University), Paolo Calafiura (Lawrence Berkeley National Lab), Taylor Childers (Argonne National Lab), Kaushik De (University of Texas at Arlington), Alessandro Di Girolamo (CERN), Esteban Fullana (University of Valencia), Wen Guan (University of Wisconsin), Tadashi Maeno (Brookhaven National Lab), Nicolo Magini (University of Genova), Paul Nilsson (Brookhaven National Lab), Danila Oleynik (University of Texas at Arlington), Shaojun Sun (University of Wisconsin), Vakho Tsulaia (Lawrence Berkeley National Lab), Peter Van Gemmeren (Argonne National Lab), Torre Wenaus (Brookhaven National Lab), and Wei Yang (Stanford Linear Accelerator Center)</i>	
Implementation of the ATLAS Trigger Within the ATLAS Multi-threaded Software Framework AthenaMT ...	339
<i>Tim Martin (University of Warwick)</i>	
Modeling Impact of Execution Strategies on Resource Utilization	340
<i>Alexey Poyda (National Research Centre "Kurchatov Institute"), Mikhail Titov (National Research Centre "Kurchatov Institute"), Alexei Klimentov (Brookhaven National Laboratory), Jack Wells (Oak Ridge National Laboratory), Sarp Oral (Oak Ridge National Laboratory), Kaushik De (University of Texas at Arlington), Danila Oleynik (University of Texas at Arlington), and Shantenu Jha (Rutgers University)</i>	
Towards Exascale Computing for High Energy Physics: The ATLAS Experience at ORNL	341
<i>V Ananthraj (OLCF), K De (University of Texas), S Jha (Rutgers University), A Klimentov (Brookhaven National Laboratory), D Oleynik (University of Texas), S Oral (OLCF), A Merzky (Rutgers University), R Mashinistov (Brookhaven National Laboratory), S Panitkin (Brookhaven National Laboratory), P Svirin (Brookhaven National Laboratory), M Turilli (Rutgers University), J Wells (OLCF), and S Wilkinson (University of Texas)</i>	
Simulating HEP Workflows on Heterogeneous Architectures	343
<i>Charles Leggett (Lawrence Berkeley National Laboratory) and Illya Shapoval (Lawrence Berkeley National Laboratory)</i>	

TrackML: A High Energy Physics Particle Tracking Challenge	344
<i>Polo Calafiura (Lawrence Berkeley National Laboratory), Steven Farrell (Lawrence Berkeley National Laboratory), Heather Gray (Lawrence Berkeley National Laboratory), Jean-Roch Vlimant (CalTech), Vincenzo Innocente (CERN), Andreas Salzburger (CERN), Sabrina Amrouche (Université de Genève), Tobias Golling (Université de Genève), Moritz Kiehn (Université de Genève), Victor Estrade (LRI), Cécile Germain (LRI), Isabelle Guyon (LRI), Ed Moyse (University of Massachusetts), David Rousseau (LAL), Yetkin Yilmaz (LAL), Vladimir Vava Gligorov (LPNHE), Mikhail Hushchyn (HSE), and Andrey Ustyuzhanin (HSE)</i>	
Automated Parallel Calculation of Collaborative Statistical Models in RooFit	345
<i>E. G. Patrick Bos (Netherlands eScience Center), Inti Pelupessy (Netherlands eScience Center), Vincent A. Croft (New York University), Wouter Verkerke (ATLAS group), and Carsten D. Burgard (ATLAS group)</i>	
Strategies for Modeling Extreme Luminosities in the CMS Simulation	347
<i>Michael D. Hildreth (University of Notre Dame), Elizabeth Sexton-Kennedy (Fermi National Accelerator Laboratory), Kevin Pedro (Fermi National Accelerator Laboratory), and Matti Kortelainen (Fermi National Accelerator Laboratory)</i>	
Deep Generative Models for Fast Shower Simulation in ATLAS	348
<i>Dalila Salamani (University of Geneva), Stefan Gadatsch (University of Geneva), Tobias Golling (University of Geneva), Graeme Andrew Stewart (CERN), Aishik Ghosh (LAL), David Rousseau (LAL), Ahmed Hasib (University of Edinburgh), and Jana Schaarschmidt (University of Washington)</i>	

Poster Session

Poster Abstracts eScience 2018 Conference	349
<i>Ben van Werkhoven (Netherlands eScience Center, The Netherlands), Adriënné Mendrik (Netherlands eScience Center, The Netherlands), and Rob van Nieuwpoort (Netherlands eScience Center, The Netherlands)</i>	
3bij3: A Framework for Testing Effects of Recommender Systems on News Exposure	350
<i>Felicia Löcherbach (Vrije Universiteit Amsterdam) and Damian Trilling (University of Amsterdam)</i>	
A First Look at the JX Workflow Language	352
<i>Tim Shaffer (University of Notre Dame), Kyle M.D. Sweeney (University of Notre Dame), Nathaniel Kremer-Herman (University of Notre Dame), and Douglas Thain (University of Notre Dame)</i>	
A Spark-Based Platform to Extract Phenological Information from Satellite Images	354
<i>Viktor Bakayov (UvA), Romulo Goncalves (NLeSC), Raul Zurita-Milla (Faculty ITC-UT), and Emma Izquierdo-Verdiguier (IPL-UValencia, Faculty ITC-UT)</i>	

Assessing and Visualising Online Information Quality	356
<i>Davide Ceolin (Information Access Group CWI), Jacco Van Ossenbruggen (Information Access Group CWI), Lora Aroyo (Vrije Universiteit Amsterdam), Ozkan Sener (Vrije Universiteit Amsterdam), Robin Sharma (Vrije Universiteit Amsterdam), Lesia Tkacz (Vrije Universiteit Amsterdam), and Julia Noordegraaf (University of Amsterdam Amsterdam)</i>	
ATLAS Trigger and Data Acquisition Upgrades for the High Luminosity LHC	358
<i>M. E. Pozo Astigarraga (CERN)</i>	
Automated Composition of Scientific Workflows in Mass Spectrometry-Based Proteomics	360
<i>Anna-Lena Lamprecht (Utrecht University), Magnus Palmblad (Leiden University Medical Center), Jon Ison (Technical University of Denmark), and Veit Schwämmle (University of Southern Denmark)</i>	
Automated Composition of Scientific Workflows: A Case Study on Geographic Data Manipulation	362
<i>Vedran Kasalica (Utrecht University) and Anna-Lena Lamprecht (Utrecht University)</i>	
Automatic Recommendations for Data Coding: A Use Case from Medical and Teacher Education	364
<i>Claudia Schulz (Technische Universität Darmstadt), Michael Sailer (LMU München), Jan Kiesewetter (LMU München), Elisabeth Bauer (LMU München), Frank Fischer (LMU München), Martin R. Fischer (LMU München), and Iryna Gurevych (Technische Universität Darmstadt)</i>	
Building the World's Largest Radio Telescope: The Square Kilometre Array Science Data Processor	366
<i>Jamie S. Farnes (University of Oxford), Ben Mort (University of Oxford), Fred Dulwich (University of Oxford), Karel Adámek (University of Oxford), Anna Brown (University of Oxford), Jan Novotny (University of Oxford), Stef Salvini (University of Oxford), and Wes Armour (University of Oxford)</i>	
Cookery: A Framework for Creating Data Processing Pipeline Using Online Services	368
<i>Mikolaj Branowski (University of Amsterdam) and Adam Belloum (University of Amsterdam)</i>	
Coordinating Asymmetries in Infrastructure Development for the Humanities	370
<i>Max Kemman (Luxembourg Centre for Contemporary and Digital History University of Luxembourg)</i>	
De-duplicating the OpenAIRE Scholarly Communication Big Graph	372
<i>Claudio Atzori (Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - CNR), Paolo Manghi (Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - CNR), and Alessia Bardi (Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - CNR)</i>	
Development of the OMUSE/AMUSE Modeling System	374
<i>Inti Pelupessy (Netherlands eScience Center), Ben van Werkhoven (Netherlands eScience Center), Gijs van den Oord (Netherlands eScience Center), Simon Portegies Zwart (Leiden Observatory), Arjen van Elteren (Leiden Observatory), and Henk Dijkstra (Institute for Marine and Atmospheric research Utrecht)</i>	
Estimating Subgraph Generation Models to Understand Large Network Formation	375
<i>Laurens Bogaardt (Netherlands eScience Center) and Frank W. Takes (University of Amsterdam)</i>	

Evaluating Layer-Wise Relevance Propagation Explainability Maps for Artificial Neural Networks	377
<i>Elena Ranguelova (Netherlands eScience Center), Eric J. Pauwels (Centrum Wiskunde & Informatica (CWI)), and Joost Berkhout (CWI)</i>	
eWaterCycle II	379
<i>Rolf Hut (Delft University of Technology), Niels Drost (Netherlands eSciencecenter), Willem Van Hage (Netherlands eSciencecenter), and Nick Van De Giesen (Delft University of Technology)</i>	
Exploiting Execution Provenance to Explain Difference Between Two Data-Intensive Computations	380
<i>Priyaa Thavasimani (Newcastle University, UK), Jacek Cala (Newcastle University, UK), and Paolo Missier (Newcastle University, UK)</i>	
Fast and Easy Mapping of Relational Data to RDF for Rapid Learning Health Care	382
<i>Martine de Vos (Netherlands eScience Center), Berend Weel (Netherlands eScience Center), Adriënne M. Mendrik (Netherlands eScience Center), Andre Dekker (Maastricht University), and Johan Van Soest (Maastricht University)</i>	
How to Bring Value of Domain Specific Big Data in an Interdisciplinary Way? A Software Landscape	384
<i>B. V. Thage (National eScience Center Danish e-infrastructure Cooperation (DeiC) Lyngby) and L. K. Andersen (National eScience Center Danish e-infrastructure Cooperation (DeiC) Lyngby)</i>	
Improving LBFGS Optimizer in PyTorch: Knowledge Transfer from Radio Interferometric Calibration to Machine Learning	386
<i>Sarod Yatawatta (ASTRON, The Netherlands Institute for Radio Astronomy), Hanno Spreeuw (Netherlands eScience Center), and Faruk Diblen (Netherlands eScience Center)</i>	
Linking Natural History Collections	388
<i>Lise Stork (Leiden University), Andreas Weber (University of Twente), Eulàlia Gassó Miracle (Naturalis Biodiversity Center Leiden), and Katherine Wolstencroft (Leiden University)</i>	
Modeling of Load Balanced Scheduling and Reliability Evaluation for On-demand Computing Based Transaction Processing System	390
<i>Dharmendra Prasad Mahato (BIT Sindri) and Jasminder Kaur Sandhu (Thapar University)</i>	
Navigating Sea-Ice Timeseries Data using Tracklines	392
<i>Brennan Bell (Jacobs University Bremen), Tilman Dinter (Alfred-Wegener-Institute), Vlad Merticariu (Jacobs University Bremen), Bang Pham Huu (Jacobs University Bremen), Dimitar Misev (Jacobs University Bremen), and Peter Baumann (Jacobs University Bremen)</i>	
Nordic Exome Variant Catalogue a Web Resource for Genomic Data Browsing	393
<i>Timo Petteri Sipilä (University of Helsinki), Andres Metspalu (University of Tartu), Paavo Häppölä (University of Helsinki), Priit Palta (University of Helsinki), Mart Kals (University of Tartu), and Aarno Palotie (University of Helsinki)</i>	
Occam: Software Environment for Creating Reproducible Research	394
<i>Luís Oliveira (University of Pittsburgh), David Wilkinson (University of Pittsburgh), Daniel Mossé (University of Pittsburgh), and Bruce R. Childers (University of Pittsburgh)</i>	

On Optimising Cost and Value in eScience: Case Studies in Radio Astronom	396
<i>P. Chris Broekema (ASTRON), Verity L. Allan (University of Cambridge), and Henri E. Bal (Vrije Universiteit Amsterdam)</i>	
Open Knowledge Discovery and Data Mining from Patient Forums	397
<i>Anne Dirkson (Leiden University), Suzan Verberne (Leiden University), Gerard Van Oortmerssen (Leiden University), Hans Gelderblom (Leiden University), and Wessel Kraaij (Leiden University)</i>	
POS: Online Learning for Memory-Aware Scheduling of Scientific Workflows	399
<i>Carl Witt (Humboldt-Universität zu Berlin), Dennis Wagner (Humboldt-Universität zu Berlin), and Ulf Leser (Humboldt-Universität zu Berlin)</i>	
Post-Processing Strategies for the ECMWF Model	401
<i>Gijs Van Den Oord (Netherlands eScience Center), Xavier Yepes (Barcelona Supercomputing Center), and Mario Acosta (Barcelona Supercomputing Center)</i>	
Scientific Partnership: A Pledge for a New Level of Collaboration between Scientists and IT Specialists	402
<i>Jens Weismüller (Leibniz Supercomputing Centre) and Anton Frank (Leibniz Supercomputing Centre)</i>	
Search for Computational Workflow Synergies in Reproducible Research Data Analyses in Particle Physics and Life Sciences	403
<i>Tibor Šimko (CERN), Kyle Cranmer (New York University), Michael R. Crusoe (Common Workflow Language), Lukas Heinrich (New York University), Anton Khodak (Wellcome Sanger Institute), Dinos Kousidis (CERN), and Diego Rodríguez (CERN)</i>	
Serving Scientists in Agri-Food Area by Virtual Research Environments	405
<i>A. Ballis (University of Athens), A. Boizet (French National Institute for Agricultural Research), L. Candela (ISTI - National Research Council of Italy), D. Castelli (ISTI - National Research Council of Italy), E. Fernández (EGI Foundation), M. Filter (Federal Institute for Risk Assessment (BfR)), T. Günther (Federal Institute for Risk Assessment (BfR)), G. Kakaletis (University of Athens), P. Karampiperis (Agroknow), D. Katris (University of Athens), M.J.R. Knapen (Wageningen University & Research), R.M. Lokers (Wageningen University & Research), L. Penev (Pensoft Publishers), G. Sipos (EGI Foundation), and P. Zervas (Agroknow)</i>	
Sight-Seeing in the Eyes of Deep Neural Networks	407
<i>Seyran Khademi (TU Delft), Xiangwei Shi (TU Delft), Tino Mager (TU Delft), Ronald Siebes (VU Universiteit Amsterdam), Carola Hein (TU Delft), Victor de Boer (VU Universiteit Amsterdam), and Jan van Gemert (TU Delft)</i>	
Smart Data Scouting in Professional Soccer: Evaluating Passing Performance Based on Position Tracking Data	409
<i>Matthias Kempe (University of Groningen, UMCG), Floris R. Goes (University of Groningen, UMCG), and Koen A. P. M. Lemmink (University of Groningen, UMCG)</i>	

SPOT: Open Source Visual Data Analytics Platform for High-Dimensional Scientific Data	411
<i>F. Diblen (Netherlands eScience Center), J.J. Attema (Netherlands eScience Center), R. Bakhshi (Netherlands eScience Center), B. Stienen (Radboud Universiteit), L. Hendriks (Radboud Universiteit), and S. Caron (Radboud Universiteit)</i>	
The Results and Challenges of Using Administrative Health Data Within a Natural Experimental Evaluation of the Abolition of Prescription Fees in Scotland	412
<i>Andrew James Williams (University of Exeter Truro), William Henley (University of Exeter), and John Frank (University of Edinburgh)</i>	
Toward VR Eventscales for Spatio-Temporal Access to Digital Maritime Heritage	413
<i>Menno Jan Kraak (University of Twente), Andreas Weber (University of Twente), Jelle van Lottum (Huygens Institute for the History of the Netherlands, KNAW), and Yuri Engelhardt (University of Twente)</i>	

Thursday 1 November: Plenary Session

DeepDownscale: A Deep Learning Strategy for High-Resolution Weather Forecast	415
<i>Eduardo Rocha Rodrigues (IBM Research), Igor Oliveira (IBM Research), Renato Cunha (IBM Research), and Marco Netto (IBM Research)</i>	
A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast	423
<i>Renato L. F. Cunha (IBM Research), Bruno Silva (IBM Research), and Marco A. S. Netto (IBM Research)</i>	
Semantic Software Metadata for Workflow Exploration and Evolution	431
<i>Lucas Augusto M. C. Carvalho (University of Campinas), Daniel Garijo (University of Southern California), Claudia Bauzer Medeiros (University of Campinas), and Yolanda Gil (University of Southern California)</i>	

Coupling Exascale Multiphysics Applications: Methods and Lessons Learned	442
<i>Jong Youl Choi (Oak Ridge National Laboratory), Choong-Seock Chang (Princeton Plasma Physics Laboratory), Julien Dominski (Princeton Plasma Physics Laboratory), Scott Klasky (Oak Ridge National Laboratory), Gabriele Merlo (The University of Texas at Austin), Eric Suchyta (Oak Ridge National Laboratory), Mark Ainsworth (Brown University), Bryce Allen (Argonne National Laboratory), Franck Cappello (Argonne National Laboratory), Michael Churchill (Princeton Plasma Physics Laboratory), Philip Davis (Rutgers University), Sheng Di (Argonne National Laboratory), Greg Eisenhauer (Georgia Institute of Technology), Stephane Ethier (Princeton Plasma Physics Laboratory), Ian Foster (Argonne National Laboratory), Berk Geveci (Kitware), Hanqi Guo (Argonne National Laboratory), Kevin Huck (University of Oregon), Frank Jenko (The University of Texas at Austin), Mark Kim (Oak Ridge National Laboratory), James Kress (Oak Ridge National Laboratory), Seung-Hoe Ku (Princeton Plasma Physics Laboratory), Qing Liu (New Jersey Institute of Technology), Jeremy Logan (University of Tennessee at Knoxville), Allen Malony (University of Oregon), Kshitij Mehta (Oak Ridge National Laboratory), Kenneth Moreland (Sandia National Laboratories), Todd Munson (Argonne National Laboratory), Manish Parashar (Rutgers University), Tom Peterka (Argonne National Laboratory), Norbert Podhorszki (Oak Ridge National Laboratory), Dave Pugmire (Oak Ridge National Laboratory), Ozan Tugluk (Brown University), Ruonan Wang (Oak Ridge National Laboratory), Ben Whitney (Brown University), Matthew Wolf (Oak Ridge National Laboratory), and Chad Wood (University of Oregon)</i>	
FDQ: Advance Analytics Over Real Scientific Array Datasets	453
<i>Roe Ebenstein (The Ohio State University), Gagan Agrawal (The Ohio State University), Jiali Wang (Argonne National Laboratory), Joshua Boley (Illinois Institute of Technology), and Rajkumar Kettimuthu (Argonne National Laboratory)</i>	
Boosting Atmospheric Dust Forecast with PyCOMPSs	464
<i>Javier Conejero (Barcelona Supercomputing Center (BSC)), Cristian Ramon-Cortes (Barcelona Supercomputing Center (BSC)), Kim Serradell (Barcelona Supercomputing Center (BSC)), and Rosa M. Badia (Barcelona Supercomputing Center (BSC), Spanish National Research Council (CSIC))</i>	
Modelling Implicit Content Networks to Track Information Propagation Across Media Sources to Analyze News Events	475
<i>Anirudh Joshi (The University of Melbourne) and Richard O. Sinnott (The University of Melbourne)</i>	
Utilizing a Transparency-Driven Environment Toward Trusted Automatic Genre Classification: A Case Study in Journalism History	486
<i>Aysenur Bilgin (CWI, Amsterdam), Erik Tjong Kim Sang (Netherlands eScience Center), Kim Smeenk (University of Groningen), Laura Hollink (CWI, Amsterdam), Jacco van Ossenbruggen (CWI, Amsterdam), Frank Harbers (University of Groningen), and Marcel Broersma (University of Groningen)</i>	
Author Index	497