

Introducing

Rescue-HPC

The 1st Workshop on Reproducible, Customizable and Portable Workflows for HPC



rescue-hpc.org

Co-designing efficient SW/HW stacks for HPC is becoming increasingly complex!

Data centers / Supercomputers

User front-end
(cloud, GRID, supercomputer, etc)

Existing frameworks / algorithms

Mobile device

Server

Algorithm / source code

Available libraries / skeletons

Compilers

Binary or byte code

Inputs

Various models

Run-time environment

Run-time state
of the system

Hardware,
simulators

XSEDE, Google Cloud, AWS, Microsoft Azure, PRACE, Watson...

Numerous models for TensorFlow, PyTorch, CNTK, MXNet

CUDA, MPI, OpenMP, TBB, OpenCL, StarPU, OmpSs ...

C, C++, Java, Python, Fortran, byte code, assembler ...

cuBLAS, BLAS, MAGMA, ViennaCL, CLBlast, cuDNN, openBLAS,
cIBLAS, libDNN, tinyDNN, ARM compute lib, libxsmm, TVM ...

LLVM, GCC, ICC, Rose, PGI, functional programming ...

diverse hardware: heterogeneous, out-of-order, caches
(CPU, GPU, GPGPU, TPU, DSP, FPGA ...)

Linux (CentOS, Ubuntu, RedHat, SUSE, Debian),
MacOS, Windows, BSD, Android ...

Continuously changing software and hardware; software package mess; lack of common experimental frameworks; no common artifact formats; Docker is not enough; little artifact reuse from papers ...

Rescue-HPC

The 1st Workshop on Reproducible, Customizable and Portable Workflows for HPC

Our aim is to bring together HPC researchers and practitioners who are interested in developing reproducible, portable and customizable experimental workflows for HPC.

We are particularly interested in contributions that describe state-of-the-art, pitfalls, comparisons and improvements to existing frameworks, benchmarks and datasets that can be used to run HPC workloads across multiple software versions and hardware architectures. Another aim of the workshop is to gather all this information in order to propose methodologies for reusing prior artifacts, automating artifact evaluation, and accelerating the development of the next generation of HPC software and hardware.

Organizers

- Grigori Fursin (dividiti, cTuning foundation)
*Author of the Collective Knowledge framework ;
SC19 reproducibility vice-chair ;*
- Todd Gamblin (LLNL)
*Author of the Spack package manager ;
SC18 program co-chair*
- Michela Taufer (University of Tennessee, Knoxville)
*SC19 general chair ;
Reproducibility champion*

Program Committee

- Lorena A Barba (George Washington University)
- Bruce Childers (University of Pittsburgh)
- Robert Clay (Sandia National Laboratories)
- Michael R. Crusoe (Common Workflow Language project)
- Kenneth Hoste (Ghent University)
- Ivo Jimenez (UC Santa Cruz)
- Daniel S Katz (NCSA)
- Arnaud Legrand (INRIA / CNRS)
- Bernd Mohr (Jülich Supercomputing Centre)
- David Richards (LLNL)
- Victoria Stodden (Stanford University)
- Flavio Vella (Free University of Bozen-Bolzano)

rescue-hpc.org

Rescue-HPC 2018 program

9:00 – 9:10	Workshop introduction
9:10 – 10:00	Keynote "Making Reproducibility Indispensable: Changing the Incentives that Drive Computational Science" <i>Michael A. Heroux, Sandia National Laboratories (SC18 reproducibility chair)</i>
10:00 – 10:30	Coffee break
10:30 – 10:45	Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context <i>Victoria Stodden and Matthew S. Krafczyk</i>
10:45 – 11:00	Considering the Development Workflow to Achieve Reproducibility with Variation <i>Michael Mercier, Adrien Faure, Olivier Richard</i>
11:00 – 11:15	Spotting Black Swans With Ease: The Case for a Practical Reproducibility Platform <i>Ivo Jimenez and Carlos Maltzahn</i>
11:15 – 11:30	Supporting Thorough Artifact Evaluation with Occam <i>Luís Oliveira, David Wilkinson, Daniel Mossé, Bruce Childers</i>
11:30 – 11:45	Semantically Organized Containers for Reproducible Research <i>Andrew Youngdahl, Zhihao Yuan, Dai-Hai Ton-That, Tanu Malik, Ivo Jimenez</i>
11:45 – 12:30	Open panel: automation, reproducibility and reusability of HPC research <i>Michela Taufer, Michael A. Heroux, Todd Gamblin, Stephen Harrell, Grigori Fursin</i>

Use the following tags for twitting: **@supercomputing** **#SC18** **#RescueHPC** **rescue-hpc.org**

Ask questions for papers: sc18.supercomputing.org/experience/sc-schedule and select a paper

Rescue-HPC 2018 keynote

Making Reproducibility Indispensable: Changing the Incentives that Drive Computational Science

Michael A. Heroux, Sandia National Laboratories (SC18 reproducibility chair)

Reproducibility is essential for scientific progress and engineering advances. Even so, many published computational results lack sufficient capture and description of companion information that would enable subsequent confirmation and extension of the results. Certainly, most scientists intend to publish correct results, but without sufficient rigor in computational processes and practices, risk is unnecessarily high that results will occasionally be wrong and will always be costly to confirm and extend.

The reasons for inadequate reproducibility are fundamentally matters of incentives and costs. In recent years, because of the availability of improved software platforms from GitHub, GitLab, Atlassian and others, and container environments such as Docker, the cost of capturing and describing the computing environment used to produce scientific results has dramatically decreased. Furthermore, new workflows and skill-building opportunities are available for those who are interested in improving their practice. What needs further attention is our incentive system.

In this presentation, we discuss efforts to improve computational reproducibility by fostering and promoting changes to our incentive systems. We talk about efforts to increase reproducibility expectations by publishers, funding agencies, employers and the broader computational scientific community. By improving incentives to produce reproducible results, providing recognition for those who lead the community and providing conduits for effective exchange of best practices, we can expect and make reproducibility indispensable.

Rescue-HPC 2018 panel

Open panel: automation, reproducibility and reusability of HPC research

- **Stephen Harrell**, Purdue University
SCC Deputy Chair & SC16 SCC Chair; SCC Emeritus
- **Michael A. Heroux**, Sandia National Laboratories
SC18 reproducibility chair
- **Michela Taufer**, University of Tennessee, Knoxville
SC19 general chair
- **Todd Gamblin**, LLNL
author of the Spack package manager ; SC18 program co-chair
- **Grigori Fursin**, dividiti/cTuning foundation
architect of the cKnowledge.org framework; SC19 reproducibility vice-chair

Three topics

- 1) major scientific and engineering challenges to reproduce results at SC and SCC;
- 2) possible engineering solutions (automated workflows, package managers, common formats and APIs for code and data sharing);
- 3) how to incentivize authors to consider reproducibility of their results and improve reusability of their code, data, models and results.

Check beta CK workflow for SCC18 (automation, reuse): <https://github.com/ctuning/ck-scc18>

Repositories of customizable, portable and reusable research components with CK API

CK JSON API

AI frameworks

TensorFlow
Caffe
Caffe2
CNTK
Torch
MXNet
...

CK JSON API

Models

AlexNet
GoogleNet
VGG
ResNet
SqueezeNet
SqueezeDet
SSD
MobileNets
...

CK JSON API

Libraries

cuDNN
ArmCL
OpenBLAS
ViennaCL
CLBlast
cuBLAS
TVM
gemmlowp
...

CK JSON API

Data sets

KITTI
COCO
VOC
ImageNet
Real life objects from the community
...

CK JSON API

Targets

Linux
MacOS
Windows
Android
...

CK JSON API

Hardware

CPU
DSP
FPGA
GPU
NN accelerators
Simulators
...

Customizable CK workflows for real-world user tasks

Assemble scenarios such as image classification as LEGO™

CK JSON API

Models

□ □ □

CK JSON API

Software

□ □ □

CK JSON API

Data sets

□ □ □

CK JSON API

Hardware

□ □ □

Share complete workflows along with published papers to automate artifact evaluation and help the community build upon prior work

Crowdsource experiments with the help of volunteers across diverse models, data sets and platforms



Present best results, workflows and components on a live scoreboard for fair comparison and reuse

cknowledge.org/repo

Help students learn multidisciplinary techniques, quickly prototype new ones, validate them in practice with companies, and even contribute back new research components

Help companies select the most appropriate workflows, save R&D costs, accelerate adoption of new techniques!