SAND2019-2224C

Enabling a Culture of Developer Productivity and Software Sustainability

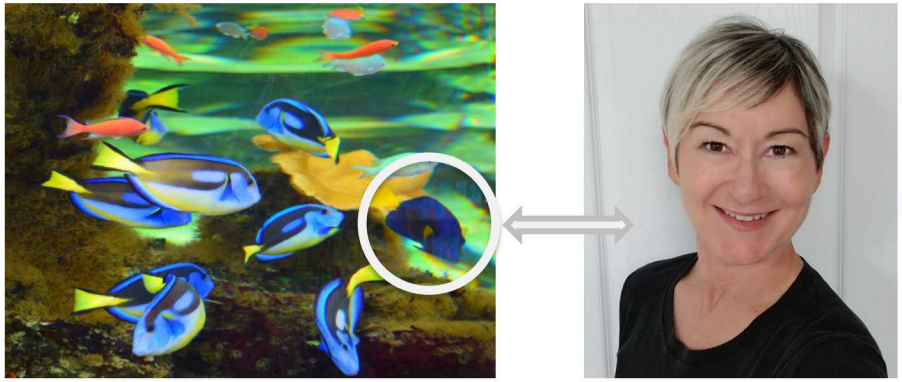
Elaine M. Raybourn
Sandia National Laboratories
emraybo@sandia.gov, @elaineraybourn

SIAM CSE March 1, 2019

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

I help organizations tell their stories.

- I am a social scientist who deepens understanding by being embedded in different cultures.



- Cultures: DARPA, DoD (Army, SOF, USMC, OSD, Team Orlando), BT Research, FhG FIT, INRIA, DOE Office of Science, Academia, National Labs
- Research: innovation and productivity, immersive learning environments, design of transmedia learning ecosystems, cultural awareness
- Focus on ECP productivity since 2017, transmedia learning since 2010, games, immersive virtual environments, social simulations, and intelligent community systems since 2000
- Passion: ***Seize opportunities that allow us to learn about ourselves and others***
- Favorite question: ***Why not?***

IDEAS productivity ECP EXASCALE COMPUTING PROJECT

IDEAS-ECP team

https://www.ideas-productivity.org

Michael Heroux (SNL), **Co-Lead PI**, **Director, Software Technology**
Lois Curfman McInnes (ANL), **Co-Lead PI**
David Bernholdt (ORNL), **Institutional PI, Outreach Lead**
Elsa Gonsiorowski (LLNL), **Institutional PI**
Osni Marques (LBNL), **Institutional PI, Webinars Lead**
David Moulton (LANL), **Institutional PI**
Boyana Norris (Univ of Oregon), **Institutional PI**
Elaine Raybourn (SNL) **Institutional PI, PSIP Lead**
Satish Balay (ANL)
Roscoe Bartlett (SNL)
Anshu Dubey (ANL)
Patricia Grubel (LANL)
Rinku Gupta (ANL), **BSSw Editor-in-Chief**
Stephen Hudson (ANL)
Reed Milewicz (SNL)
Mark Miller (LLNL)
Jared O'Neal (ANL)
Barry Smith (ANL)
Greg Watson (ORNL)
Jim Willenbring (SNL), **SDK Lead**
Paul Wolfenbarger (SNL)
Lisa Childers (ALCF)
Rebecca Hartman-Baker (NERSC)
Judy Hill (OLCF)
Hai Ah Nam (LANL), **BSSw Fellows**
Jean Shuler (LLNL)

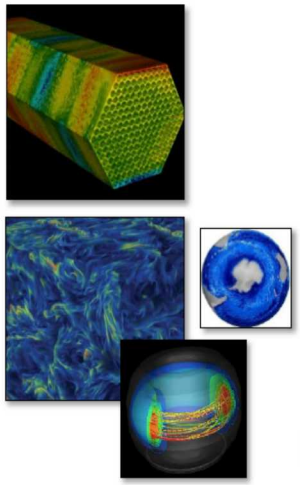
Computing Facilities Liaisons

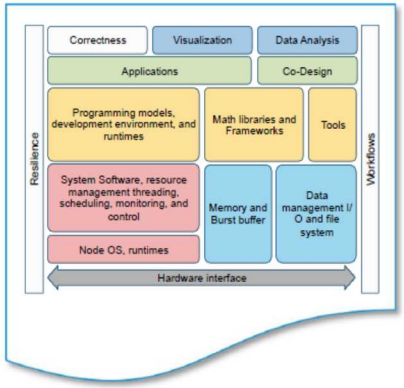
What is the Exascale Computing Project?

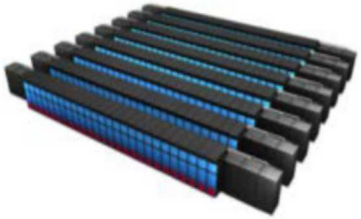
- As part of the National Strategic Computing initiative, ECP was established to accelerate delivery of a **capable exascale computing system** that integrates hardware and software capability to deliver approximately 50 to 100 times more performance than today's petaflop machines.
- ECP's work encompasses applications, system software, hardware technologies and architectures, and workforce development to meet the scientific and national security mission needs of DOE.

To achieve capable exascale requires a holistic approach

Application Development	Software Technology	Hardware Technology	Exascale Systems
Science and mission applications	Scalable and productive software stack	Hardware technology elements	Integrated exascale supercomputers







ECP's work encompasses applications, system software, hardware technologies and architectures, and workforce development

IDEAS productivity

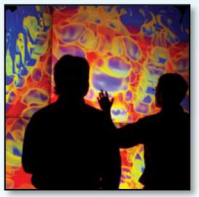
ECP EXASCALE COMPUTING PROJECT

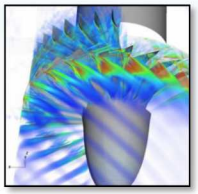
ECP applications target six strategic areas

National security

Stockpile stewardship

Next-generation electromagnetics simulation of hostile environment and virtual flight testing for hypersonic re-entry vehicles





Energy security

Turbine wind plant efficiency

High-efficiency, low-emission combustion engine and gas turbine design

Materials design for extreme environments of nuclear fission and fusion reactors

Design and commercialization of Small Modular Reactors

Subsurface use for carbon capture, petroleum extraction, waste disposal

Scale-up of clean fossil fuel combustion

Biofuel catalyst design

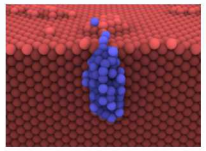
Economic security

Additive manufacturing of qualifiable metal parts

Reliable and efficient planning of the power grid

Seismic hazard risk assessment

Urban planning



Scientific discovery

Find, predict, and control materials and properties

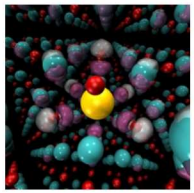
Cosmological probe of the standard model of particle physics

Validate fundamental laws of nature

Demystify origin of chemical elements

Light source-enabled analysis of protein and molecular structure and design

Whole-device model of magnetically confined fusion plasmas



Earth system

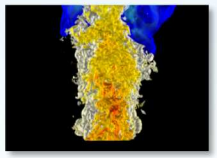
Accurate regional impact assessments in Earth system models

Stress-resistant crop analysis and catalytic conversion of biomass-derived alcohols

Metagenomics for analysis of biogeochemical cycles, climate change, environmental remediation

Health care

Accelerate and translate cancer research



IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

ECP Goals

- **Application Development**: Deliver a broad array of comprehensive science-based computational applications that effectively utilize exascale HPC technology to provide breakthrough simulation and data analytic solutions for scientific discovery, energy assurance, economic competitiveness, health enhancement, and national security
- **Ease of Use**: Create software that makes exascale systems usable by a wide variety of scientists and engineers across a range of applications
- **Diverse Architectures**: Enable by 2023 ≥ two diverse computing platforms with up to 50× more computational capability than today's 20 PF systems, within a similar size, cost, and power footprint
- **US HPC Leadership**: Help ensure continued American leadership in architecture, software and applications to support scientific discovery, energy assurance, stockpile stewardship, and nonproliferation programs and policies

ECP by the Numbers

A seven-year, $1.7 B R&D effort that launched in 2016

Six core DOE National Laboratories: Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, Sandia

- Staff from most of the 17 DOE national laboratories take part in the project

Three technical focus areas (Application Development, Software Technology, Hardware and Integration)

More than 100 top-notch R&D teams

- Hundreds of consequential milestones delivered on schedule and within budget since project inception

IDEAS productivity ECP EXASCALE COMPUTING PROJECT

Many ECP ST products are available (many github)

For example…

Challenges of CSE for ECP – *so many stories!*

Technical

- All parts of the ecosystem can be under research
- Requirements change throughout the lifecycle as knowledge grows
- Importance of reproducibility, sustainability
- Verification is complicated
- Real world is messy, so is the software

Social

- Competing priorities and incentives
- Limited resources
- Perception of “invisible work” with deferred or no benefit
- Need for interdisciplinary interactions
- Boutique operations must scale!

Science through computing is only as good as the software that produces it.

IDEAS-ECP enables a culture of developer productivity

- Deliver value to ECP application teams and software technologies. This overarching goal drives all of our work.
- Engagement with the broader software community -- adapting and adopting approaches, and raising awareness of the particular needs of extreme-scale computational science.
- Development and dissemination of better scientific software practices
- Incorporation of informed strategies for human systems: focusing on changing the way scientific teams and individuals work.
- Engagement with leadership computing facilities: direct liaison roles for project leaders at all major DOE computing facilities.
- Web-based content development and delivery: expanding the development and usability of the Better Scientific Software web portal (https://bssw.io) as a go-to site for content on developer productivity and software sustainability.

IDEAS-ECP Project

https://www.ideas-productivity.org, https://bssw.io

Co-PIs: Lois McInnes (ANL) & Michael Heroux (SNL)

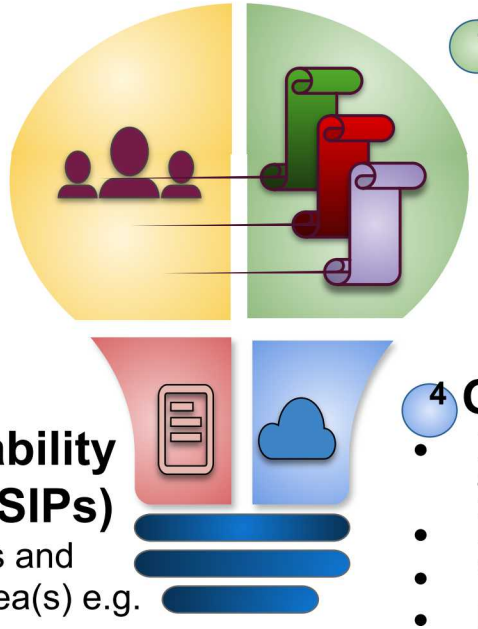
Goal: Improve Exascale Computing Project (ECP) developer productivity and software sustainability while ensuring continued scientific success.

1 Interviews with Exascale Computing teams

- Applications & Software Technology
- Understand crosscutting productivity challenges, priorities, and opportunities

2 Productivity and Sustainability Improvement Planning (PSIPs)

- Work with team to define focus and track progress on particular area(s) e.g. research software engineering



3 Customize, create, and curate methodologies

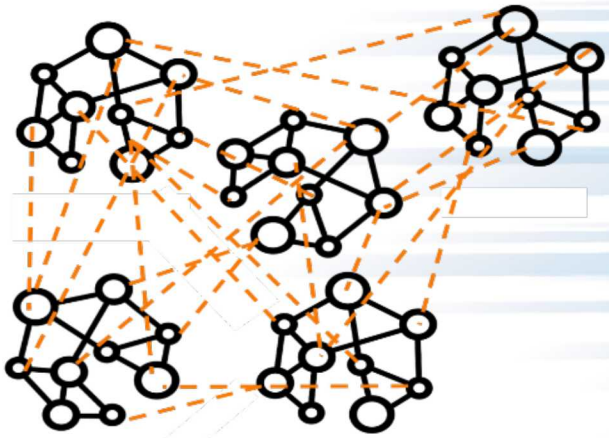
- Targeting application productivity and sustainability
- Create user stories to convey requirements from interview & PSIPs to determine priorities, plans for work

4 Outreach and training

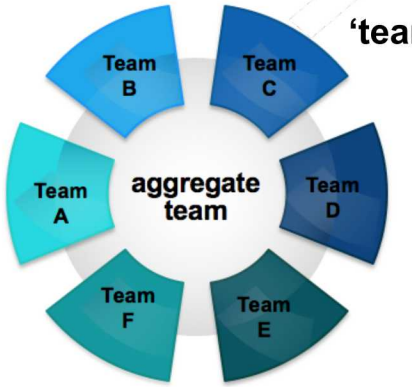
- In partnership with US Department of Energy facilities
- Documents: WhatIs, HowTo, PSIP policies
- Webinar series and tutorials
- Better Scientific Software site (https://bssw.io)

Interactions with ECP teams have illuminated needs

- Testing/verification of scientific software
- Team onboarding and team member transitions
- Intermediate/advanced Git (especially for aggregate teams)
- Agile team management
- Agile workflows for scientific software
- Use of (interoperable) scientific libraries



‘teams of teams”



Process for interviews, synthesis, outreach

ECP Application, Co-Design, and Software Teams:
CANDLE, ExaGraph, Exascale MPI, ExaStar, E3SM-MMF, EXAALT, MARBL, NWChemEx, UnifyCR, QMCPack, and WDMApp

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

HI Training and Productivity: Integration with ST and AD

Members of the ECP Productivity Project (i.e., IDEAS-ECP) work with ST teams to:

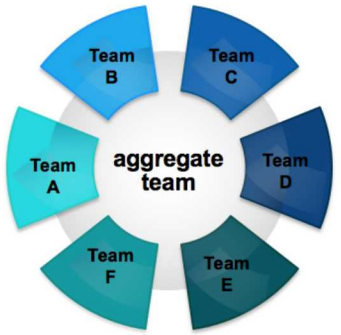
- **Understand** current software practices
- **Identify** crosscutting, high-priority needs for training and outreach
- **Collaborate** on Productivity and Sustainability Improvement Planning (PSIP)
- **Improve** software practices while maintaining scientific productivity; share best practices experiences

Process for interviews, synthesis, outreach

ECP Application, Co-Design, and Software Teams:
CANDLE, ExaGraph, Exascale MPI, ExaStar, E3SM-MMF, EXAALT, MARBL, NWChemEx, UnifyCR, QMCPack, and WDMApp

High-priority needs for ECP teams:
Tools that foster productive and sustainable collaboration (through software) for ***aggregate*** ECP science teams

- testing/verification of scientific software
- team onboarding and team member transitions
- intermediate/advanced Git (especially for aggregate teams)
- code reviews for identifying defects
- agile team management,
- agile workflows for scientific software
- use of (interoperable) scientific libraries



Many ECP projects are ***aggregate*** *teams*, composed of multiple successful previously existing teams, where software is a primary vehicle of collaboration.

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

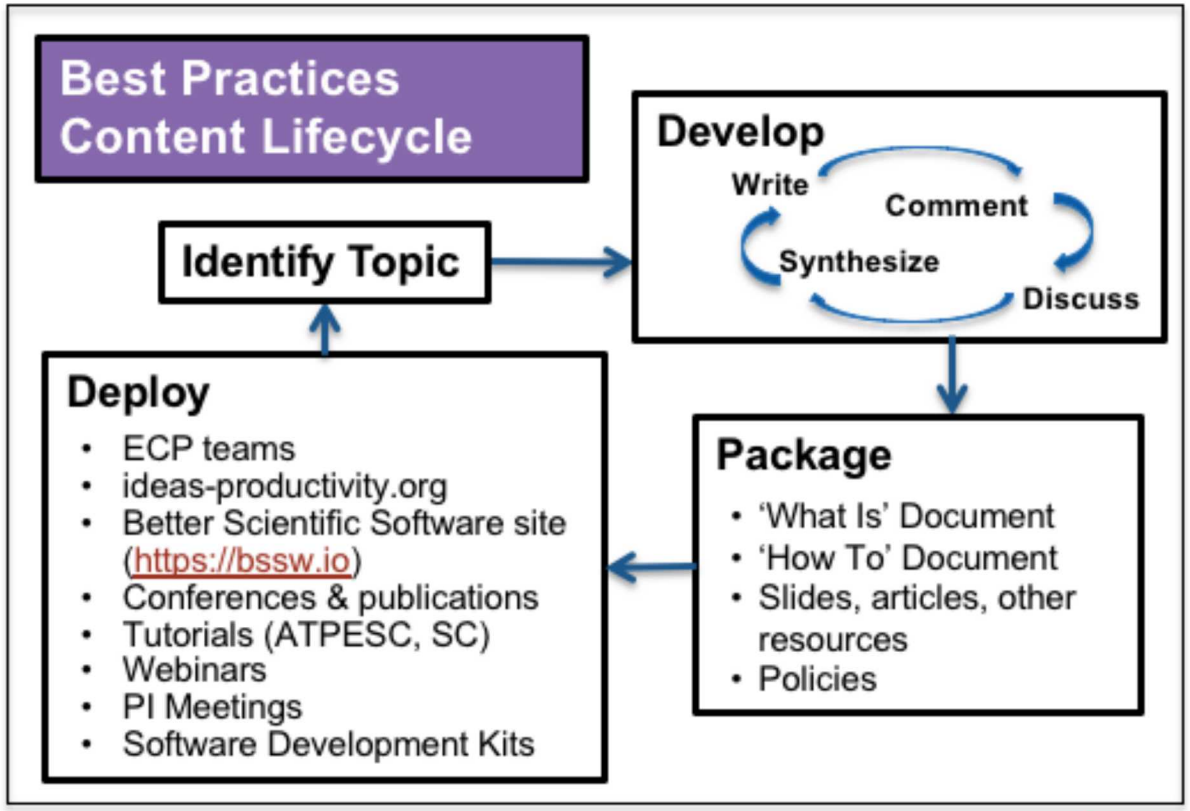
Workflow for Best Practices Content Development

Modern learning theory:

Build from knowledge base: Elaboration and models

Vast body of SE content from broad community

Learn, adapt, adopt, assimilate



Productivity and Sustainability Improvement Planning (PSIP) Examples: EXAALT & MPICH

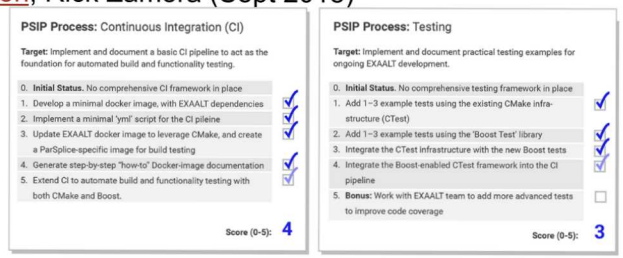
PSIP workflow helps a team create user stories, identify areas for improvement, select a specific area and topic for a single improvement cycle, and then develop those improvements with specific metrics for success.

MPICH PSIP: Onboarding new team members

Practice: Create Centralized Training Resources		
Score (0 – 4)	Description	Tracking
0	Initial Status : No training process in place.	
1	Understand MPICH requirement for developers and typical challenges for new hires	✓
2	Review and gather specific training materials	✓
3	Design "MPICH Training Base" website	✓
4	Solicit feedback, improve, add and prune content to ensure effectiveness	2019

EXAALT PSIP: Continuous integration (CI) testing

BSSw blog article:
Adopting Continuous Integration for Long Timescale Materials Simulation, Rick Zamora (Sept 2018)



PSIP Process: Continuous Integration (CI)

Target: Implement and document a basic CI pipeline to act as the foundation for automated build and functionality testing.

0. **Initial Status.** No comprehensive CI framework in place
1. Develop a minimal docker image, with EXAALT dependencies ✓
2. Implement a minimal 'yml' script for the CI pileine ✓
3. Update EXAALT docker image to leverage CMake, and create a ParSplice-specific image for build testing ✓
4. Generate step-by-step "how-to" Docker-image documentation ✓
5. Extend CI to automate build and functionality testing with both CMake and Boost. ✓

Score (0-5): 4

PSIP Process: Testing

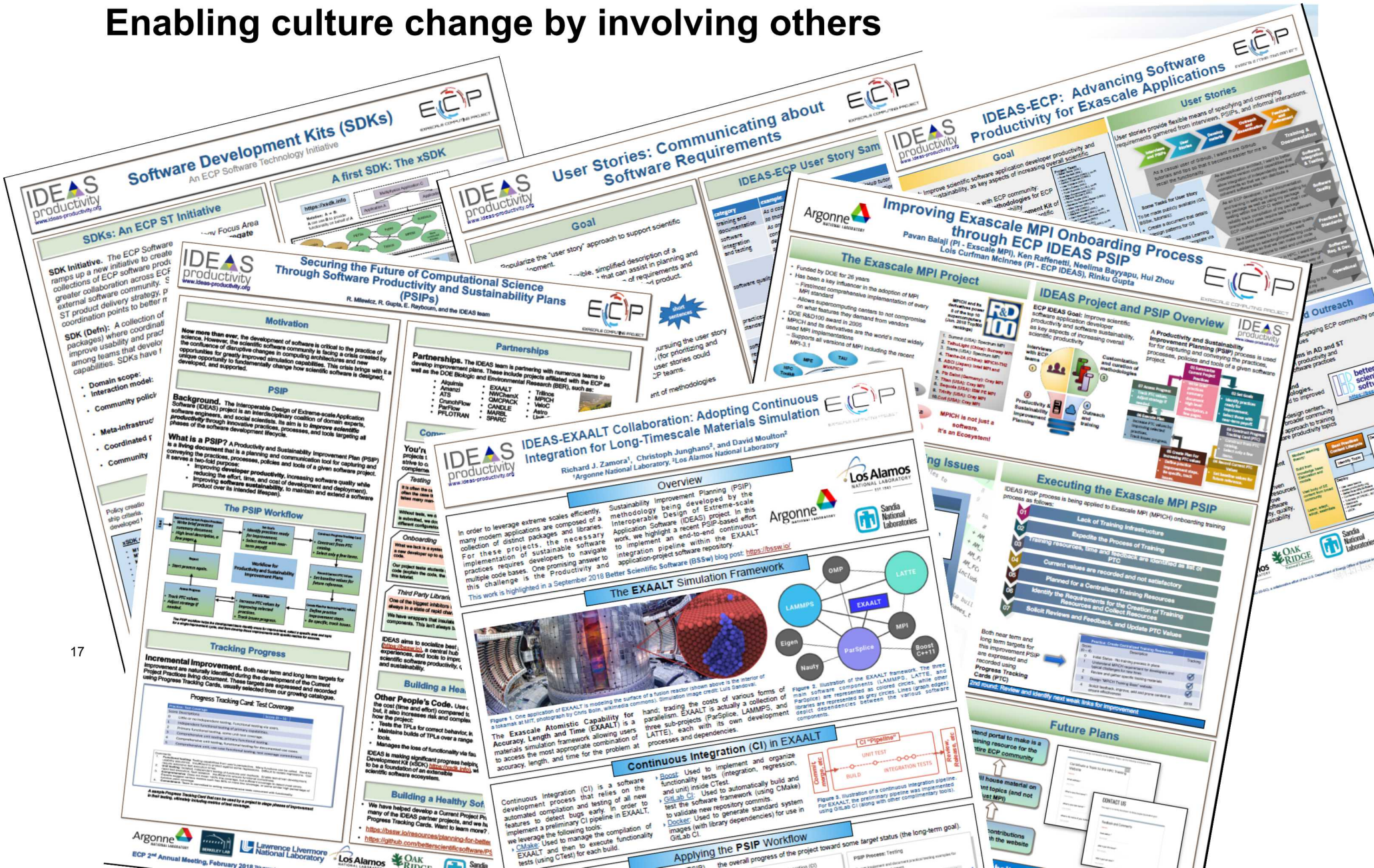
Target: Implement and document practical testing examples for ongoing EXAALT development.

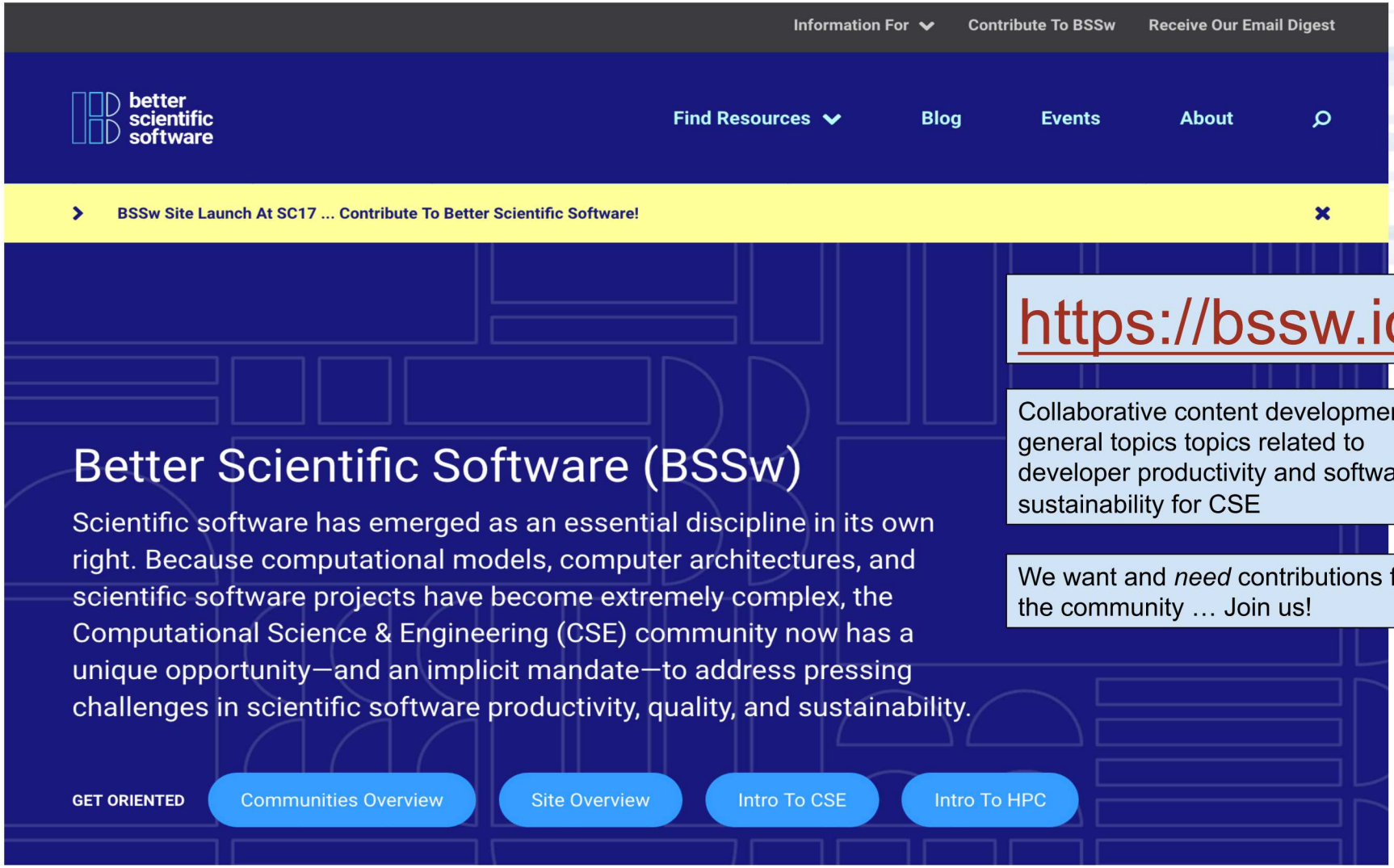
0. **Initial Status.** No comprehensive testing framework in place ✓
1. Add 1–3 example tests using the existing CMake infra-structure (CTest) ✓
2. Add 1–3 example tests using the 'Boost Test' library ✓
3. Integrate the CTest infrastructure with the new Boost tests ✓
4. Integrate the Boost-enabled CTest framework into the CI pipeline ✓
5. **Bonus:** Work with EXAALT team to add more advanced tests to improve code coverage ☐

Score (0-5): 3

IDEAS productivity ECP EXASCALE COMPUTING PROJECT

Enabling culture change by involving others



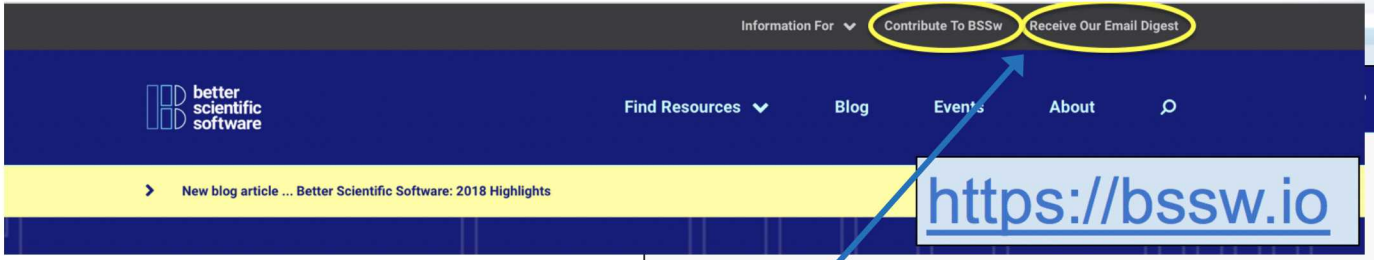


https://bssw.io

Collaborative content development on general topics topics related to developer productivity and software sustainability for CSE

We want and *need* contributions from the community … Join us!

better scientific software



What is BSSw?

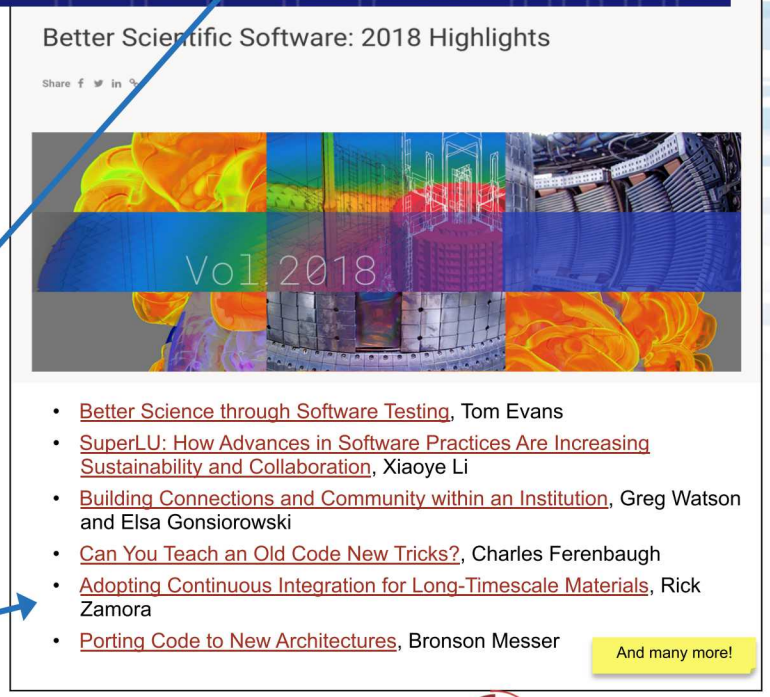
Community-based resource for sharing information on practices, techniques, and tools to improve developer productivity and software sustainability for computational science and engineering.

We want and *need* contributions from the community … Join us!

- **Types of content**
  - Informative articles
  - Curated links
    - Highlight other web-based content
  - Events
  - WhatIs, HowTo docs
  - Blog articles

Receive our email digest

Many ECP contributors

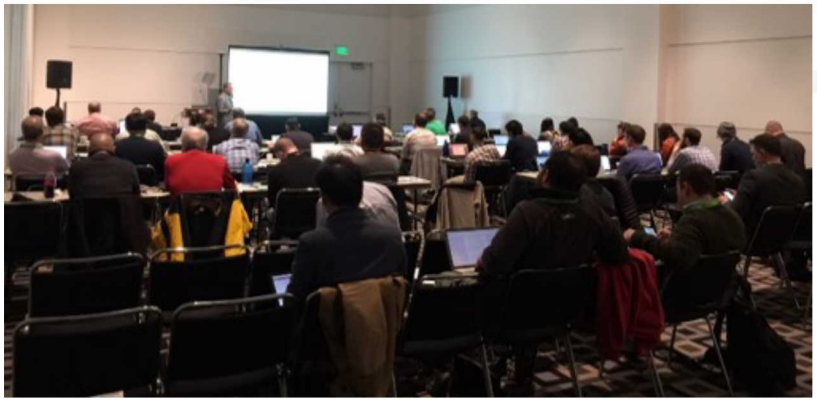


IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

HPC Best Practices Webinar Series

- 2017-06-07 Python in HPC, Rollin Thomas (NERSC), William Scullin (ALCF) and Matt Belhorn (OLCF)
- 2017-09-13 Barely Sufficient Project Management: A few techniques for improving your scientific software development, Mike Heroux (SNL)
- 2017-11-01 Managing Defects in HPC Software Development, Tom Evans (ORNL)
- 2018-01-17 Bringing Best Practices to a Long-Lived Production Code Charles Ferenbaugh (LANL)
- 2018-04-18 Software Citation Today and Tomorrow, Daniel Katz (NCSA and UIUC)
- 2018-06-13 Popper: Creating Reproducible Computational and Data Science Experimentation Pipelines, Ivo Jimenez (UCSC)
- 2018-07-18 How Open Source Software Supports the Largest Computers on the Planet, Ian Lee (LLNL)
- 2018-08-21 Software Sustainability: Lessons learned from different disciplines, Neil Chue Hong (Software Sustainability Institute, UK)
- 2018-09-19 Modern CMake, Bill Hoffman (Kitware)



IDEAS productivity ECP EXASCALE COMPUTING PROJECT

better scientific software

BSSw Fellowship Program

Recognition & funding to leaders and advocates of high-quality scientific software

We will begin accepting applications for the 2020 BSSw Fellowship Program in September, 2019

We are looking for people who are:

- Passionate about scientific software.
- Interested in contributing powerful ideas, tools, methodologies, and more that improve the quality of scientific software.
- Able to use the fellowship to broadly benefit the scientific software community.
- Willing to participate as an alum in subsequent years to guide selection of future fellows and promote better scientific software in their community.

https://bssw.io *So your code will see the future*

Subscribe to BSSw mailing list to be notified about BSSw Fellowship for 2020

IDEAS productivity ECP EXASCALE COMPUTING PROJECT

License, citation and acknowledgements

License and Citation

- Requested citation: Raybourn, E.M. Enabling a Culture of Developer Productivity and Software Sustainability. 2019 SIAM Conference on Computational Science and Engineering, Spokane, WA, 2019. March 1, 2019. SAND2019-XXXX C. DOI: https://doi.org/XXXXX.

Acknowledgements

- Special thanks to the members of IDEAS-ECP.
- This work was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. Images used by permission.

IDEAS productivity ECP EXASCALE COMPUTING PROJECT