



# Measuring software productivity and sustainability?

David E. Bernholdt (he/him)  
Oak Ridge National Laboratory

Panel: Methods, Challenges, and Opportunities in Measuring the Impact of CI Projects, Metrics 2023

learn more about IDEAS at <https://ideas-productivity.org>  
and <https://doi.org/10.48550/arXiv.2311.02010>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

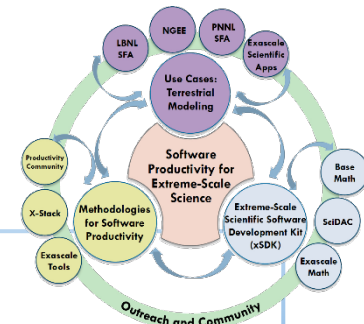
*This work was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration*

# A Brief History of IDEAS

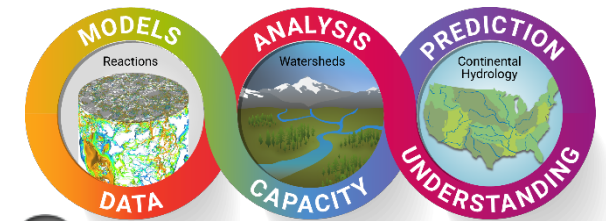
- IDEAS = Interoperable Design of Extreme-Scale Application Software
- First of its kind (in U.S.) with a focus on incubating, curating, and disseminating knowledge and methodologies about the sustainment of scientific software
  - Inspired by UK [Software Sustainability Institute](#)
- IDEAS is now a family of related projects
  - Different sponsors, different time frames, different people (but significant overlap), different approaches
  - **Common focus on improving developer productivity and software sustainability and trustworthiness**

## Acronyms

DOE	= U.S. Department of Energy
ASCR	= Office of Advanced Scientific Computing Research
BER	= Office of Biological and Environmental Research
ECP	= Exascale Computing Project



- [IDEAS-Classic](#) (2014—2017)
  - Focus: multiscale multiphysics terrestrial ecosystem modeling
  - Sponsors: DOE/ASCR and BER
- [IDEAS-ECP](#) (2017—2023)
  - Focus: supporting the ecosystem of applications, libraries, and tools developed by ECP
  - Sponsor: DOE/ECP
- [IDEAS-Watersheds](#) (2019—present)
  - Focus: accelerating watershed science through a community driven software ecosystem
  - Sponsor: DOE/BER



IDEAS  
productivity

ECP  
EXASCALE  
COMPUTING  
PROJECT

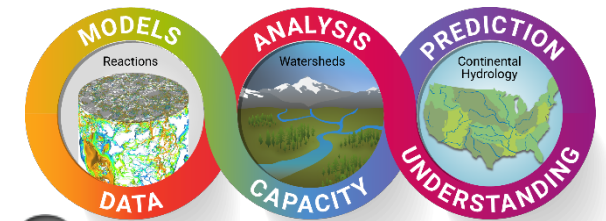
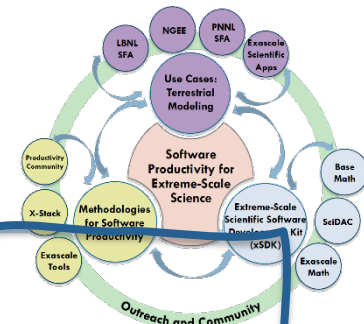
# A Brief History of IDEAS

- IDEAS = Interoperable Design of Extreme-Scale Application Software
- First of its kind (in U.S.) with a focus on incubating, curating, and disseminating knowledge and methodologies about the sustainment of scientific software
  - Inspired by UK [Software Sustainability Institute](#)
- IDEAS is now a family of related projects
  - Different sponsors, different time frames, different people (but significant overlap), different approaches
  - **Common focus on improving developer productivity and software sustainability and trustworthiness**

## Acronyms

DOE	= U.S. Department of Energy
ASCR	= Office of Advanced Scientific Computing Research
BER	= Office of Biological and Environmental Research
ECP	= Exascale Computing Project

- [IDEAS-Classic](#) (2014—2017)
  - Focus: multiscale multiphysics terrestrial ecosystem modeling
  - Sponsors: DOE/ASCR and BER
- [IDEAS-ECP](#) (2017—2023)
  - Focus: supporting the ecosystem of applications, libraries, and tools developed by ECP
  - Sponsor: DOE/ECP
- [IDEAS-Watersheds](#) (2019—present)
  - Focus: accelerating watershed science through a community driven software ecosystem
  - Sponsor: DOE/BER



IDEAS  
productivity

ECP  
EXASCALE  
COMPUTING  
PROJECT

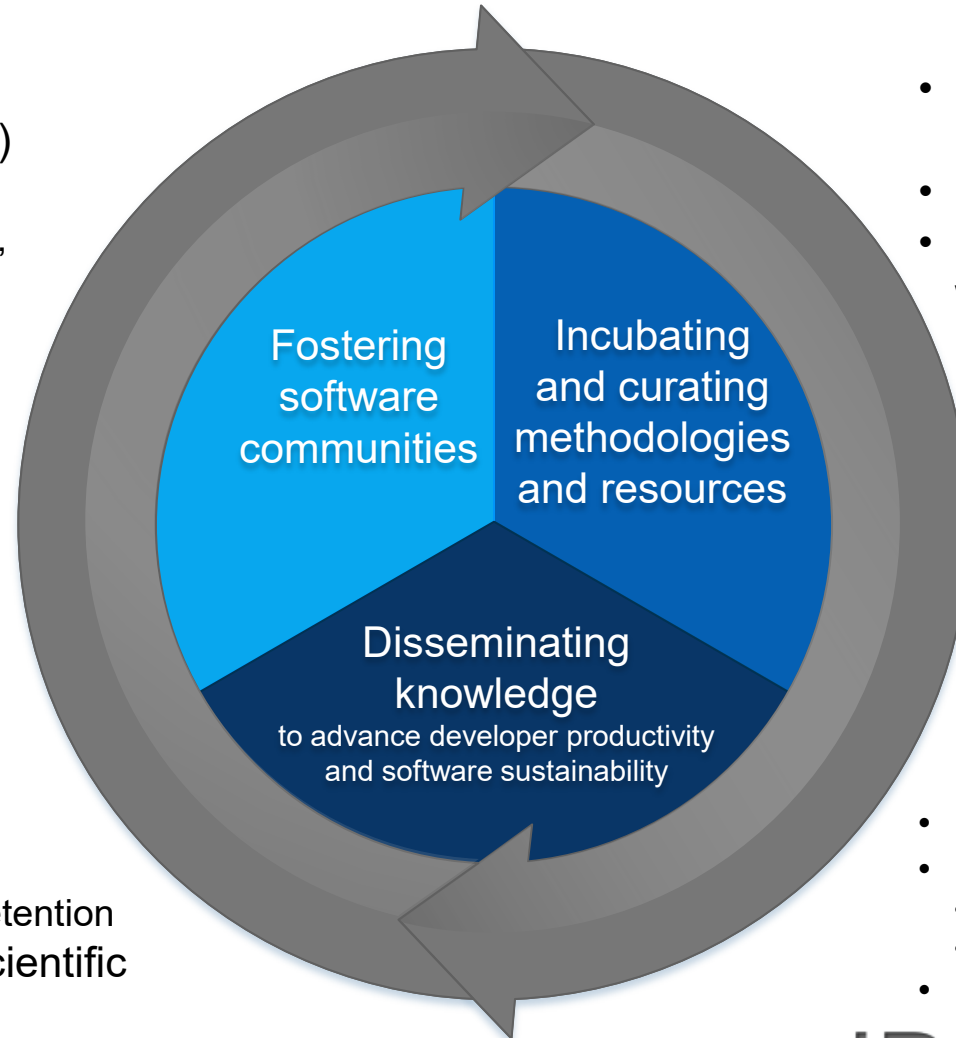
# Quantifying Software Productivity and Sustainability?

- No singular definition of either “software productivity” or “software sustainability”
  - Variations in what they mean to each practitioner
- Nevertheless, there have been many attempts to quantify
  - None are entirely satisfactory
- Mike Heroux (SNL): the better approach is to think about an eye exam – you can tell which of two options is better
  - It is an individual thing
- Early on, we attempted to use the Goal-Questions-Metrics methodology
  - Never got much that we felt was useful
  - We were probably doing it wrong



# So What Are we Doing in IDEAS?

- Software community policies
- Software Development Kits (SDKs) and E4S
  - xSDK, CAT-SDK, DAV-SDK, SWAS, etc



- Productivity and Sustainability Improvement Planning (PSIP)
- *Team of teams* concepts
- Better Scientific Software (BSSw.io) website

- Webinar Series:
  - HPC Best Practices
  - HPC Workforce Development and Retention
- Tutorials on Practices for Better Scientific Software

- BSSw Fellowship Program
- Panel Series:
  - Strategies for Working Remotely
  - Performance Portability
- Events: BOFs, workshops, and more





# xSDK: Primary delivery mechanism for ECP math libraries' continual advancements

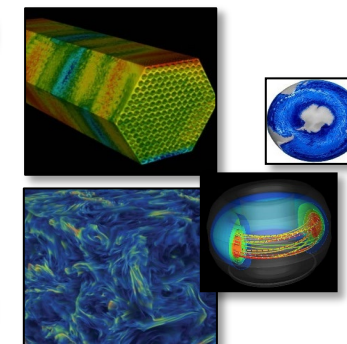
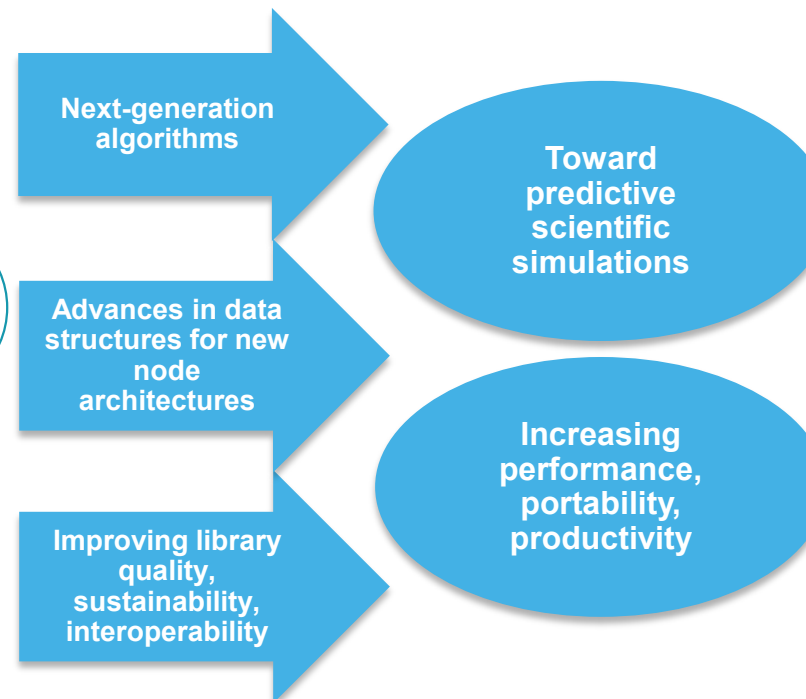
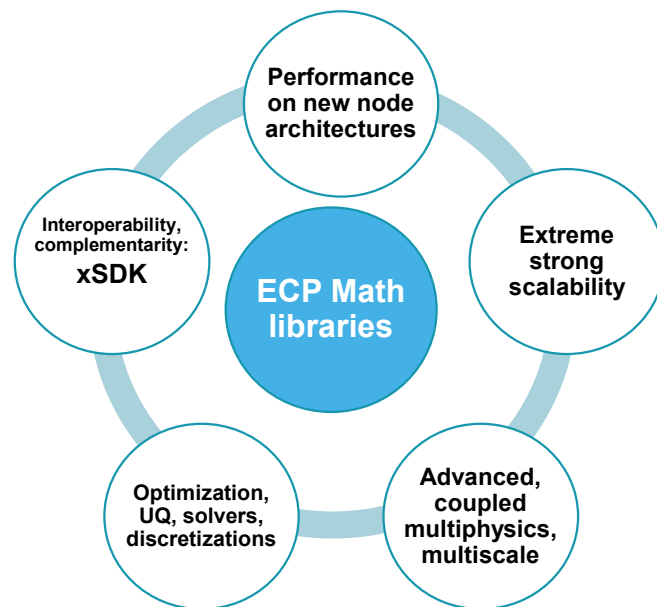


xSDK lead: Ulrike Meier Yang (LLNL)  
xSDK release lead: Satish Balay (ANL)

## xSDK release 0.8.0 (Nov 2022)

hypre	} the initial xSDK 0.1.0 libraries
PETSc/TAO	
SuperLU	
Trilinos	
Alquimia	} domain libraries included in xSDK 0.1.0 or 0.2.0
PFLOTRAN	
AMReX	
ArborX	
ButterflyPACK	
DTK	
Ginkgo	
heFFTe	
libEnsemble	
MAGMA	
MFEM	
Omega_h	
PLASMA	
PUMI	
SLATE	
Tasmanian	
SUNDIALS	
Strumpack	
deal.II	} from the broader community
preCICE	
PHIST	
SLEPc	

As motivated and validated by the needs of ECP applications:



Timeline:



Refs: [xSDK: Building an Ecosystem of Highly Efficient Math Libraries for Exascale](#), **SIAM News**, Jan 2021; [Building Community through xSDK Software Policies](#), HPC-BP webinar, Dec 2019

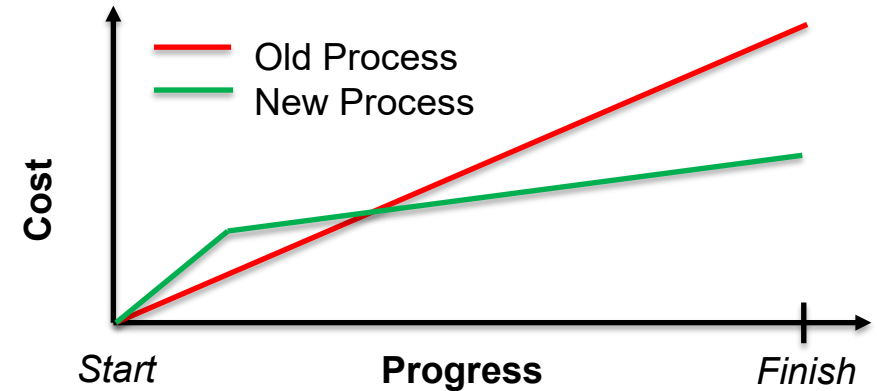


# PSIP: Productivity and Sustainability Improvement Planning

## Continual, Incremental Software Process Improvement

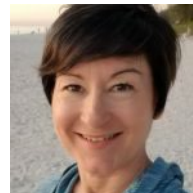
<https://bssw.io/psip>

1. Identify your team's "pain points" in your software development processes
    - Help: RateYourProject assessment tool:  
<https://rateyourproject.org/>
  2. Set a goal for something to improve
    - Target processes and behaviors, not just tasks
    - Pick something that you can address in a few months that will give you a noticeable benefit
  3. Agree on a plan to address it, identify markers of progress and what is "done"
    - Write them down
    - Help: Progress tracking card examples:  
<https://bssw-psip.github.io/ptc-catalog/catalog>
  4. Work your plan, track your progress
  5. When you are done, celebrate...
- ...then pick a new pain point to address



*The new process costs something to implement, but it pays off over time*

Target: your project should include "just enough" software engineering so that you can meet your short-term and longer-term scientific goals effectively



Lead: Elaine  
Raybourn (SNL)

# Examples of Who's using PSIP



Improvements to documentation, setting code style standards, transition to GitHub ([blog article](#))

"The PSIP project had an immediate impact on our community. With the GitHub move we see increasing amounts of small but very valuable contributions to make HDF5 code and documentation better." – **Elena Pourmal, Director of Engineering, The HDF Group**



Improve testing and verification, transition development workflow to GitHub

Revamp build system, implement a CTest-based testing framework, implement a basic CI pipeline



ALPINE/ZFP

Created a VTK-m filter for APLINE in situ algorithm users



Using a more detailed version for internal project assessment

Completed PSIP tutorial, investigating how it can be used in academic context



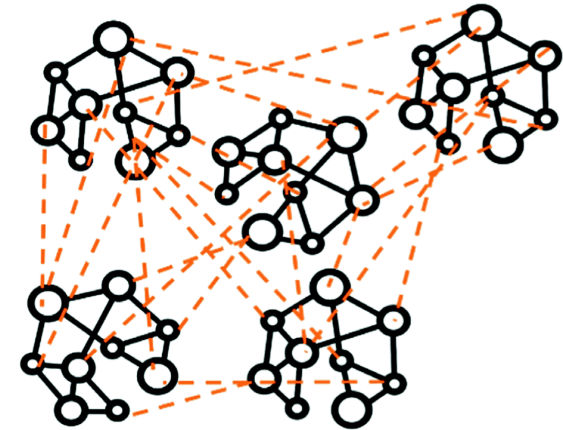
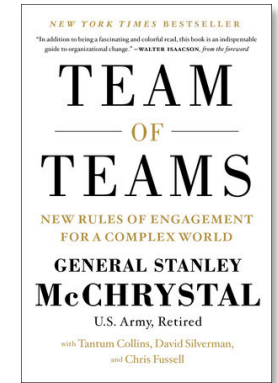
Using internally for reproducibility LDRD research, and for large projects updating version control systems, and updating documentation to support better onboarding





# Collaboration via Teams of Teams

- The “team of teams” concept (ToT) was popularized by Stanley McChrystal’s 2015 book
  - IDEAS efforts are an offshoot of PSIP, led by Elaine Raybourn (SNL)
  - Using tools from the CAT-SDK software community for repository analysis
- ToT provides a powerful lens through which to better understand the ECP, as well as many other software ecosystems, and to improve their effectiveness
  - Strengthen community partnerships
  - Scaling productivity typically experienced in small teams (where it's easy), to larger groups via the team of teams paradigm
- ToT principles facilitated contributions of the HDF5 team to the E4S and Data & Viz SDK
  - Supported applications in modeling earthquakes, electronic structures, subsurface flow, reacting flow, stellar explosions, wind plants, and cosmology
- Distributed, Interconnected Teams through the Lens of Team of Teams Principles
  - Panel discussion with members of PETSc, Trilinos, xSDK, and E4S ECP projects
- Scaling productivity and innovation on the path to exascale with a “team of teams” approach
  - Case study of the ASC Ristra ECP project



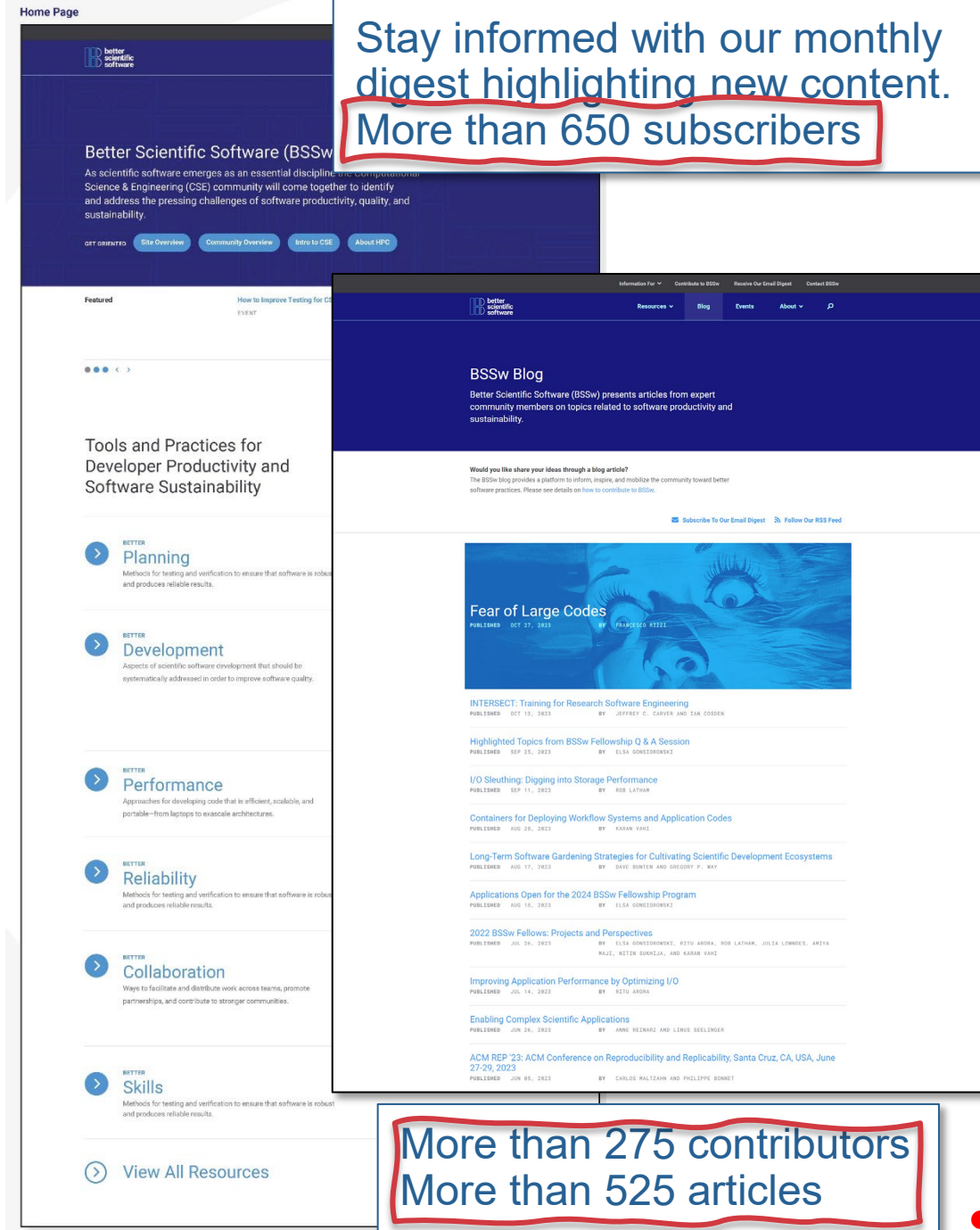
Schematic illustration  
of a team of teams,  
from doi:[10.1007/978-3-030-22338-0\\_33](https://doi.org/10.1007/978-3-030-22338-0_33)

A central hub for sharing information on practices, techniques, experiences, and tools to improve developer productivity and software sustainability for computational science & engineering (CSE)

- **Find information** on scientific software topics
- **Contribute new resources** based on your experiences
- Editor-in-chief: Rinku Gupta (ANL)

## Types of content on BSSw

- **Blog articles:** success stories, perspectives, opportunities, technical deep-dives, and more
- **Curated content:** short pointers to useful material already hosted elsewhere
- **Events:** increase awareness of events related to better scientific software



Stay informed with our monthly digest highlighting new content.  
More than 650 subscribers

Better Scientific Software (BSSw) is a central hub for sharing information on practices, techniques, experiences, and tools to improve developer productivity and software sustainability for computational science & engineering (CSE).

BSSw Blog  
Better Scientific Software (BSSw) presents articles from expert community members on topics related to software productivity and sustainability.

Would you like share your ideas through a blog article?  
The BSSw blog provides a platform to inform, inspire, and mobilize the community toward better software practices. Please see details on [how to contribute to BSSw](#).

Subscribe To Our Email Digest   Follow Our RSS Feed

**Fear of Large Codes**  
PUBLISHED OCT 27, 2023 BY FRANCISCO RIZZI

**INTERSECT: Training for Research Software Engineering**  
PUBLISHED OCT 15, 2023 BY JEFFREY C. CARVER AND SAN CODDEN

**Highlighted Topics from BSSw Fellowship Q & A Session**  
PUBLISHED SEP 25, 2023 BY ELISA GONZIMORSKI

**I/O Sleuthing: Digging into Storage Performance**  
PUBLISHED SEP 11, 2023 BY ROB LATHAM

**Containers for Deploying Workflow Systems and Application Codes**  
PUBLISHED AUG 28, 2023 BY KARAN VAID

**Long-Term Software Gardening Strategies for Cultivating Scientific Development Ecosystems**  
PUBLISHED AUG 17, 2023 BY DAVE BUNTER AND GREGORY P. BAY

**Applications Open for the 2024 BSSw Fellowship Program**  
PUBLISHED AUG 15, 2023 BY ELISA GONZIMORSKI

**2022 BSSw Fellows: Projects and Perspectives**  
PUBLISHED JUN 26, 2023 BY ELISA GONZIMORSKI, RITU ARORA, ROB LATHAM, JULIA LOWMEYER, ARVIND MAJJI, NEITAN SUDRIZLA, AND KARAN VAID

**Improving Application Performance by Optimizing I/O**  
PUBLISHED JUN 14, 2023 BY RITU ARORA

**Enabling Complex Scientific Applications**  
PUBLISHED JUN 26, 2023 BY ANNE HEINRICH AND LINDS SEELINGER

**ACM REP '23: ACM Conference on Reproducibility and Replicability, Santa Cruz, CA, USA, June 27-29, 2023**  
PUBLISHED JUN 05, 2023 BY CARLOS KALTZAHN AND PHILIPPE BONNET

More than 275 contributors  
More than 525 articles

# Better Scientific Software (BSSw) Fellowship Program



## Meet Our Fellows

The BSSw Fellowship program gives recognition and funding to leaders and advocates of high-quality scientific software. Meet the Fellows and Honorable Mentions and learn more about how they impact Better Scientific Software.

Fellowships Overview


Apply

Meet Our Fellows


BSSw Fellowship FAQ

Recognizing leaders  
2018 - 2023


**2018 Class**  
Fellows




**Jeffrey Carver**  
University of Alabama  
Improving code quality through modern peer code review



**Ivo Jimenez**  
University of California, Santa Cruz  
Enabling reproducible research through automated computational experimentation




**Daniel S. Katz**  
University of Illinois at Chicago  
Supercomputing Applications




**Andrew Lumsdaine**  
Pacific Northwest National Laboratory  
Enabling reproducible research through automated computational experimentation


**Honorable Mentions**




**Neal Davis**  
University of Illinois at Urbana-Champaign  
Teaching Assistant Professor, Computer Science



**Marc Henry de Frahan**  
National Renewable Energy Laboratory  
Postdoctoral Researcher



**Elsa Gonsiorowski**  
Lawrence Livermore National Laboratory  
HPC VSI Specialist, Livermore Computing



**Ying Li**  
Argonne National Laboratory  
Argonne Science, Argonne Leadership Computing Facility

**2019 Class**  
Fellows



**Rene Gassmoeller**  
University of California, Davis  
Building your scientific software project from inception to long-term sustainability



**Ignacio Laguna**  
Lawrence Livermore National Laboratory  
Improving the reliability of scientific applications by enabling and debugging floating point software



**Tanu Malik**  
DePaul University  
Reducing the critical debt in scientific software through reproducible containers



**Kyle Niemeyer**  
Oregon State University  
Educating scientists on best practices for developing research software

**Honorable Mentions**



**Stephen Andrews**  
Los Alamos National Laboratory  
Staff Scientist, SCHE Verification and Analysis



**Nasir Eddy**  
University of Alabama  
Ph.D. Student, Computer Science




**Benjamin Pritchard**  
Virginia Tech  
Software Scientist, Molecular Science Software Institute




**Vanessa Schacht**  
Stanford University  
Research Software Engineer, Stanford Research Computing Center


**2020 Class**  
Fellows



**Nasir Eddy**  
University of Alabama  
Automating testing in scientific software




**Damian Reardon**  
Sustainable Horizons Institute, Storrery Institute  
Responsible scientific software development in underrepresented groups




**Cindy Rubio-Gonzalez**  
University of California, Davis  
Improving the reliability and performance of numerical software


**Honorable Mentions**



**David Boehme**  
Lawrence Livermore National Laboratory  
Research Staff, Center for Applied Scientific Computing




**Sumanta Harshadewar**  
Chargent Consulting  
Founder and President, Open source software management and collaboration




**David Rogers**  
National Center for Computational Sciences, Oak Ridge National Lab  
Computational Scientist


**2021 Class**  
Fellows




**Marisol Garcia-Rojas**  
Purdue Institute  
Increasing accessibility of data & cloud technologies



**Mary Ann Leung**  
Sustainable Horizons Institute  
Increasing developer productivity and innovation through diversity




**Chase Millon**  
Mission Concepts  
Project management best practices for research software




**Amy Roberts**  
University of Colorado Denver  
Enabling collaboration through virtual assistant user stories


**Honorable Mentions**




**Keith Beattie**  
Lawrence Berkeley National Laboratory  
Computational Research Scientist, Computer Systems Engineer



**Julia Stewart**  
National Center for Ecological Analysis and Synthesis (NCEAS), US Santa Barbara  
OpenAccess Director




**Jonathan Madson**  
National Laboratory NERSC, Application Performance Specialist




**Addi Thakur**  
Malviya  
Software Engineering Group, Group Leader


**2022 Class**  
Fellows




**Ritu Arora**  
University of Texas at San Antonio  
Optimizing I/O for better performance




**Rob Latham**  
Argonne National Laboratory  
I/O enabling a new set of challenges and solutions




**Julia Stewart**  
Lawrence Livermore National Laboratory  
Improving security, privacy, package management



**Amiya K. Maji**  
Purdue University  
Securing scientific software development




**Nitin Sukhija**  
Shriya Sukhija University of Pennsylvania  
Secure scientific software development




**Karan Vahi**  
USC Information Sciences Institute  
Scientific workflows for high efficiency HPC


**Honorable Mentions**




**Sarah Brett**  
Argonne Computing  
Ph.D. Student, School of Information Studies




**William Godoy**  
Oak Ridge National Laboratory  
Computer Science




**Brittany Johnson**  
Mazda  
Assistant Professor, Computer Science Department



**Meghan Jones**  
University of Illinois at Urbana-Champaign  
Postdoctoral Researcher, Department of Earth Sciences, School of Space & Earth Sciences & Technology



**Rafael Mudafort**  
National Renewable Energy Laboratory  
Research Software Engineer, National Renewable Energy Center



**Quanghui Wu**  
University of Tennessee, Knoxville  
Assistant Professor, Department of Geography

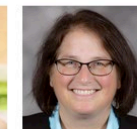
## 2023 Class

Fellows



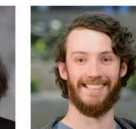
**Nicole Brewer**  
Arizona State University

Improving accessibility of data and software with scientific web apps



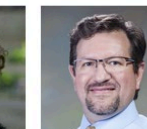
**Myra Cohen**  
Iowa State University

Techniques for scientific software testing



**Johannes Doerfert**  
Lawrence Livermore National Laboratory

Demystifying the compiler black box



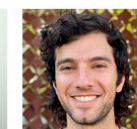
**William Hart**  
Sandia National Laboratories

Best practices for software supply chain security



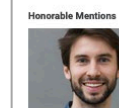
**Helen Kershaw**  
National Center for Atmospheric Research

Improving code review skills for scientific software developers



**Rafael Mudafort**  
National Renewable Energy Laboratory

Effective communication of software design



**Jean Luca Bez**  
Lawrence Berkeley National Laboratory

Scientific Data Division, Postdoctoral Researcher



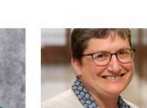
**Jose Monsalve Diaz**  
Argonne National Laboratory

Postdoctoral Researcher, Mathematics & Computer Science Division



**Xu Liu**  
North Carolina State University

Associate Professor, Computer Science



**Alisa Neeman**  
Muskingum University

Assistant Professor, Mathematics and Computer Science



**Kristina Riemer**  
University of Arizona

Scientific Programmer, Data Science Institute



**Brigitta Sipöcz**  
California Institute of Technology

Applications Developer

**Goal: Foster and promote practices, processes, and tools to improve developer productivity and software sustainability of scientific codes.** #somycodewillseethefuture

Also supported by the National Science Foundation since 2021



BSSw Fellowship Coordinator:  
Elsa Gonsiorowski (LLNL)



Deputy Coordinator, Community  
Building: Erik Palmer (LBNL)



# Better Scientific Software Tutorials

- Covering issues of developer productivity, software sustainability and reliability, with a special focus on the challenges of complex, large-scale HPC
  - software design, agile methodologies, Git workflows, reproducibility, software testing, continuous integration testing, refactoring, and more

• <https://bssw-tutorial.github.io>

• Lead: David Bernholdt (ORNL)

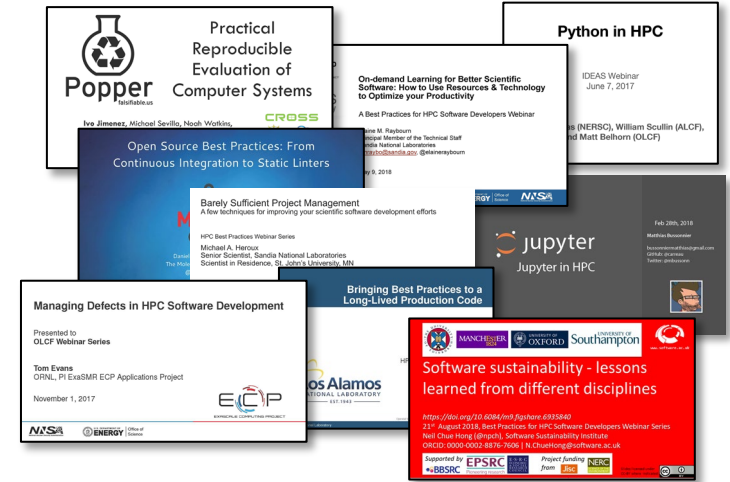
• 32 tutorials since 2016

• Presentations (all) and recordings (some) available

• Topics and content under continuous refinement

• Frequent venues

- Supercomputing (2016-2023)
- ISC (2018-2019, 2021-2023)
- ATPESC (2016-2023)



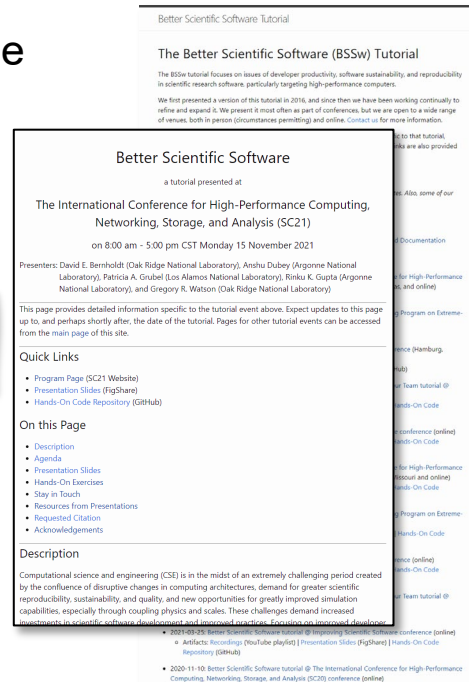
## Webinar Series: Best Practices for HPC Software Developers (HPC-BP)

- Covering topics in software development and HPC
- <https://ideas-productivity.org/resources/series/hpc-best-practices-webinars/>
- Lead: Osni Marques (LBNL)
- Presented by the community to the community
- Monthly series, since May 2016 (offered live and archived)

But < 100 responses to post-webinar feedback survey!

To date: 80 webinars, >12,000 registrations, >5,300 attendees

84 attendees per webinar, on average



# Panel Series: Performance Portability & ECP

- Lead: Anshu Dubey (ANL). Refs:
  - [Performance Portability in the Exascale Computing Project: Exploration Through a Panel Series](#), A. Dubey et al, IEEE CiSE, Sept 2021
  - SIAM CSE21 minisymposium: <https://doi.org/10.6084/m9.figshare.c.5321441>
  - ECCOMAS 2022 minisymposium

No data because there was no registration, no recording, and no one specifically noting headcounts during events



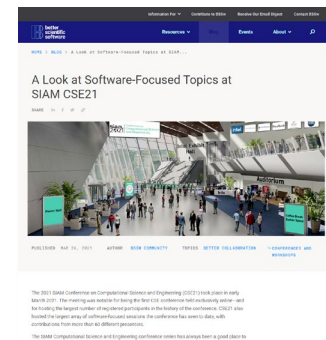
# Panel Series: Strategies for Working Remotely

- Exploring strategies for working remotely, with emphasis on how HPC teams can be effective and efficient in long-term hybrid settings
- <https://www.exascaleproject.org/strategies-for-working-remotely>
- Lead: Elaine Raybourn (SNL)
- Quarterly series, since April 2020 (offered live and archived)
- Ref: [Why We Need Strategies for Working Remotely: The ECP Panel Series](#), E. Raybourn, SC20 State of the Practice, Nov 2020

Data comparable to HPC-BP webinars

# Technical Meetings and Birds of a Feather Sessions

- Creating opportunities to talk about software development, productivity, and sustainability
- <https://ideas-productivity.org/resources/series/technical-sessions-and-meetings/>
- Minisymposia
  - SIAM CSE, SIAM PP (2015-2023), PASC (2018, 2019)
  - Ref: [A Look at Software-Focused Topics at SIAM CSE21](#), March 2021
- Thematic poster sessions
  - SIAM CSE (2017, 2019, 2021)
- BOF sessions
  - Software Engineering and Reuse in Modeling, Simulation and Data Analytics for Science and Engineering
    - <http://bit.ly/swe-cse-bof>
    - Supercomputing (2015-2023), ISC (2019, 2022-2023)
- [Collegeville Workshop Series on Scientific Software](#),
  - Ref: [Software Team Experiences and Challenges](#), K. Beattie et al, Oct 2021



Limited data – if we remember to count heads, try to generate artifacts



# IDEAS-ECP Impacts – Survey Results

- Feedback underscores IDEAS's role in **enhancing software quality, promoting best practices, and expanding awareness** of the importance of software development
- Curating best practices for software development and team productivity has **empowered teams** to build new practices into their workflows and increase cross-project collaboration
  - Many community members express a desire for additional resources
- Software communities have proven to be a **source of inspiration** for building shared foundations for software ecosystems while respecting team autonomy
- IDEAS outreach mechanisms have enabled innovators in scientific software practices to **share knowledge** with the community
- *Model for other multi-institutional software ecosystems*

IRB approval for surveys is painful, response rates are low



# Summary

- IDEAS is *not* a traditional research project, and it is *not* intended to produce software. IDEAS is intended to help **other people** do a **better job stewarding their software**
- Software productivity and sustainability are not uniquely defined, nor are there satisfactory metrics
  - Optometrist metric: is A better or is B better? Only individual can judge
- Collecting feedback is hard
  - Managed to get a few testimonial quotes, survey with low response rate
- So we ended up counting things
  - Customer engagements (SDKs, PSIP), events organized (tutorials, webinars, technical sessions, etc.), registrations and attendees, technical session speakers, etc.
  - In the “heat of battle” we often neglect to do audience counts
- And we try to leave artifacts behind
  - Presentation slides, event recordings, blog articles, etc.