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 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

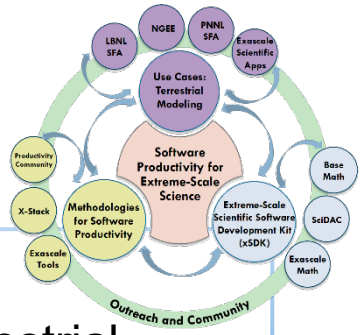
U.S. DEPARTMENT OF ENERGY | Office of Science
NNSA National Nuclear Security Administration

A Brief History of IDEAS

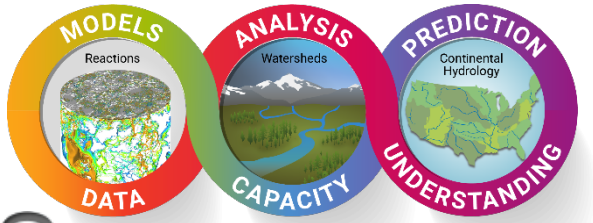
- IDEAS = Interoperable Design of Extreme-Scale Application Software
- First of its kind (in U.S.) with a focus on incubating, curating, and disseminating knowledge and methodologies about the sustainment of scientific software
  - Inspired by UK Software Sustainability Institute
- IDEAS is now a family 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

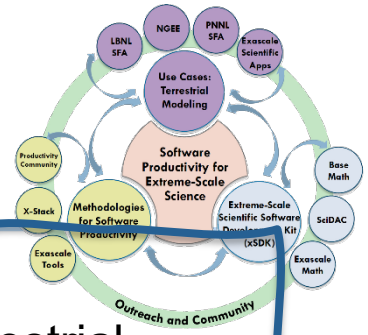
ECP EXASCALE COMPUTING PROJECT

A Brief History of IDEAS

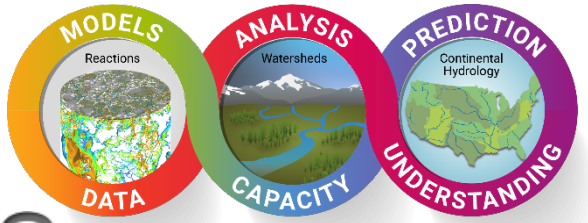
- IDEAS = Interoperable Design of Extreme-Scale Application Software
- First of its kind (in U.S.) with a focus on incubating, curating, and disseminating knowledge and methodologies about the sustainment of scientific software
  - Inspired by UK Software Sustainability Institute
- IDEAS is now a family 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?

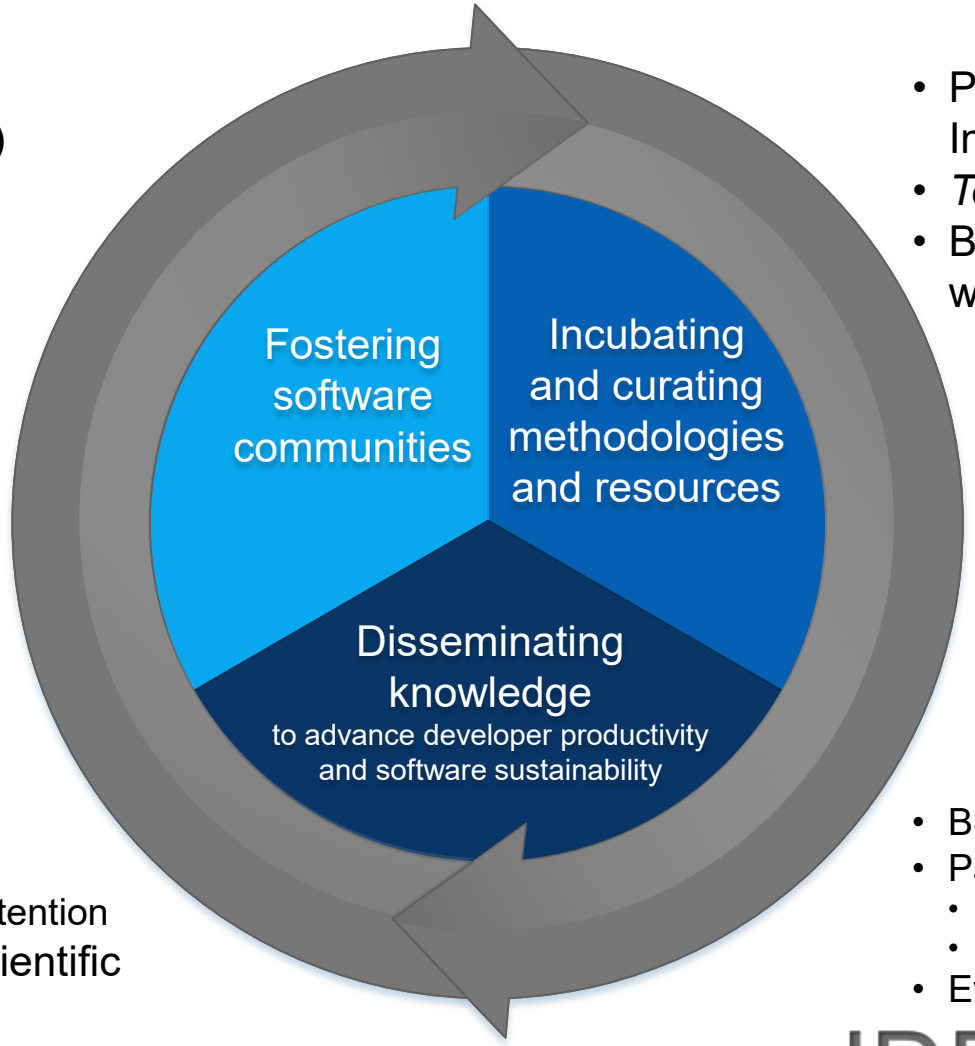
- No singular definition of either “software productivity” or “software sustainability”
  - Variations in what they mean to each practitioner
- Nevertheless, there have been many attempts to quantify
  - None are entirely satisfactory
- Mike Heroux (SNL): the better approach is to think about an eye 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 metholdology
  - Never got much that we felt was useful
  - We were probably doing it wrong

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

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

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

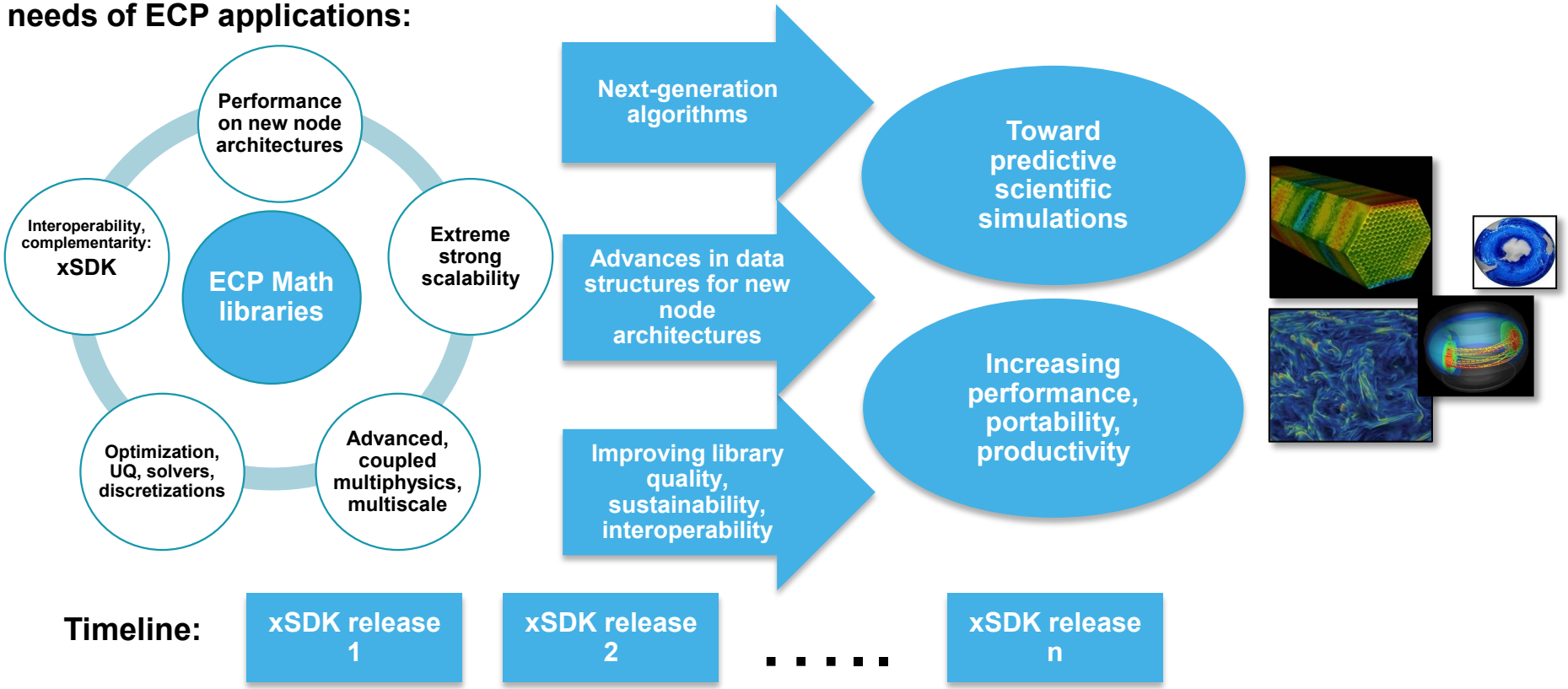
xSDK: Primary delivery mechanism for ECP math libraries' continual advancements


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

xSDK release 0.8.0 (Nov 2022)

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

As motivated and validated by the needs of ECP applications:



Timeline: xSDK release 1 → xSDK release 2 → . . . . . → xSDK release n

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

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

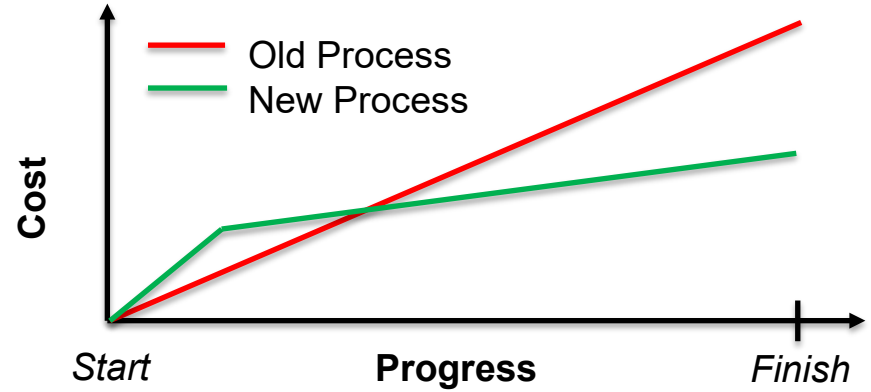
PSIP: Productivity and Sustainability Improvement Planning
Continual, Incremental Software Process Improvement

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

Lead: Elaine Raybourn (SNL)



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

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

Examples of Who's using PSIP

> "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**

The HDF Group

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

FLASH5

Improve testing and verification, transition development workflow to GitHub

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

EXAALT

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

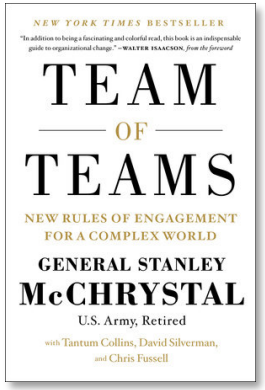
Using internally for reproducibility LDRD research, and for large projects updating version control systems, and updating documentation to support better onboarding

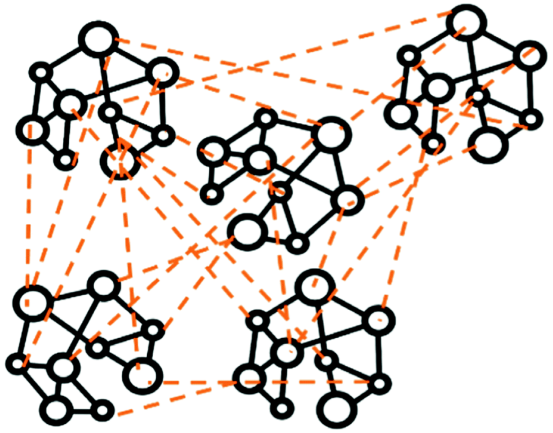
IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

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

better scientific software

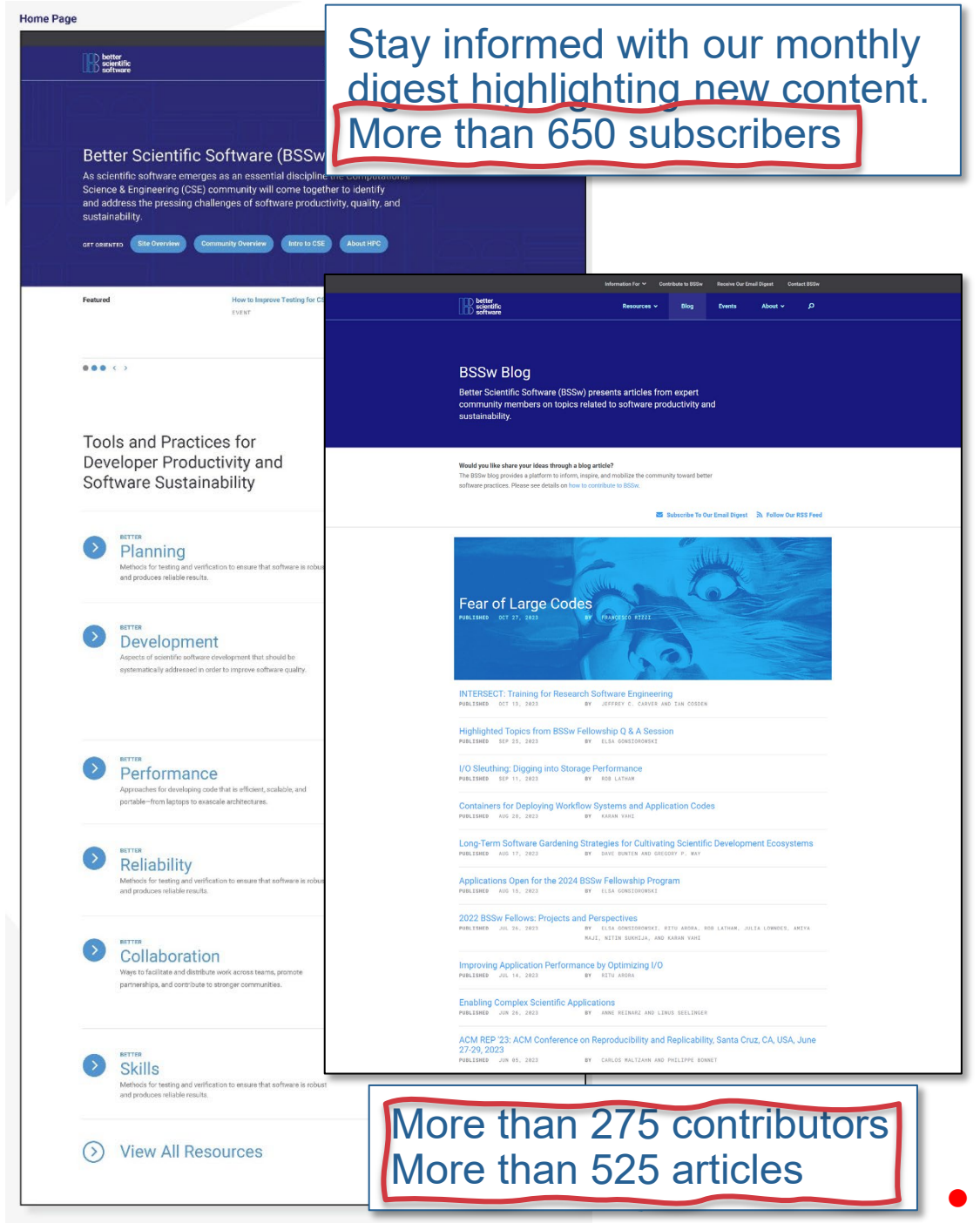
https://bssw.io

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

More than 275 contributors
More than 525 articles

Better Scientific Software (BSSw) Fellowship Program

better scientific software

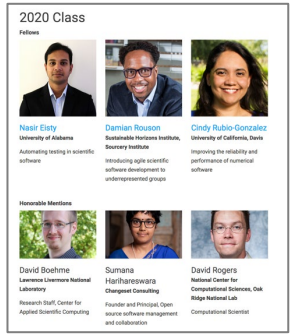
Meet Our Fellows

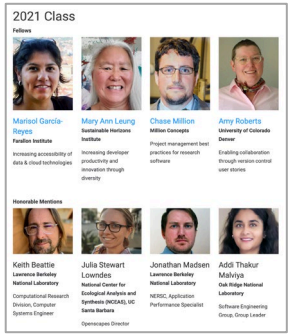
The BSSw Fellowship program gives recognition and funding to leaders and advocates of high-quality scientific software. Meet the Fellows and Honorable Mentions and learn more about how they impact Better Scientific Software.

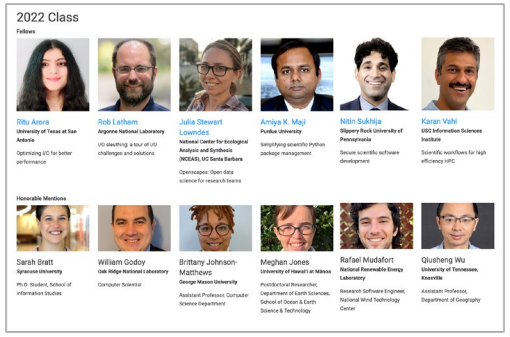
Fellowships Overview | Apply | Meet Our Fellows | BSSw Fellowship FAQ

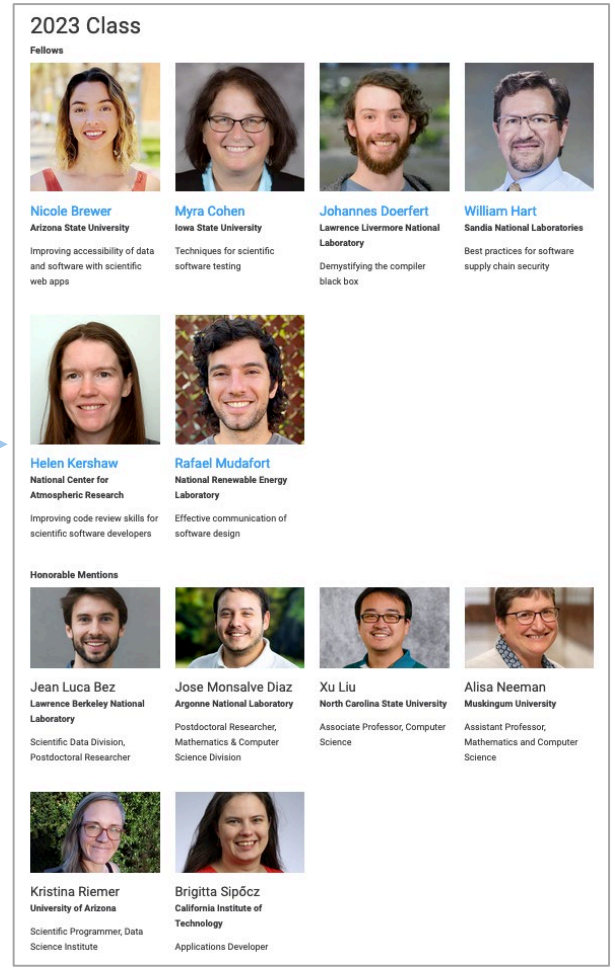
Recognizing leaders

2018 - 2023









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

BSSw Fellowship Coordinator: Elsa Gonsiorowski (LLNL)

Deputy Coordinator, Community Building: Erik Palmer (LBNL)

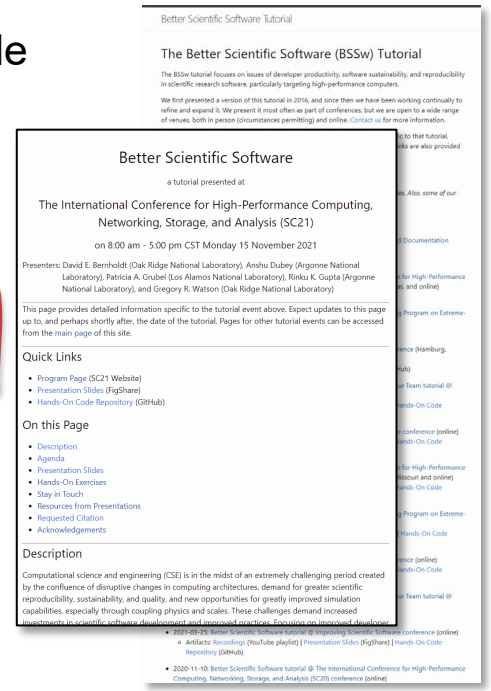
Also supported by the National Science Foundation since 2021

IDEAS productivity

ECP EXASCALE COMPUTING PROJECT

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)
  - To date: 80 webinars, >12,000 registrations, >5,300 attendees
  - 84 attendees per webinar, on average

But < 100 responses to post-webinar feedback survey!

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

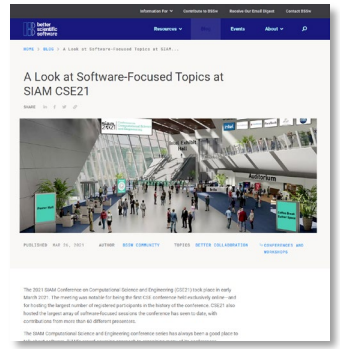
Panel Series: Strategies for Working Remotely

- 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.