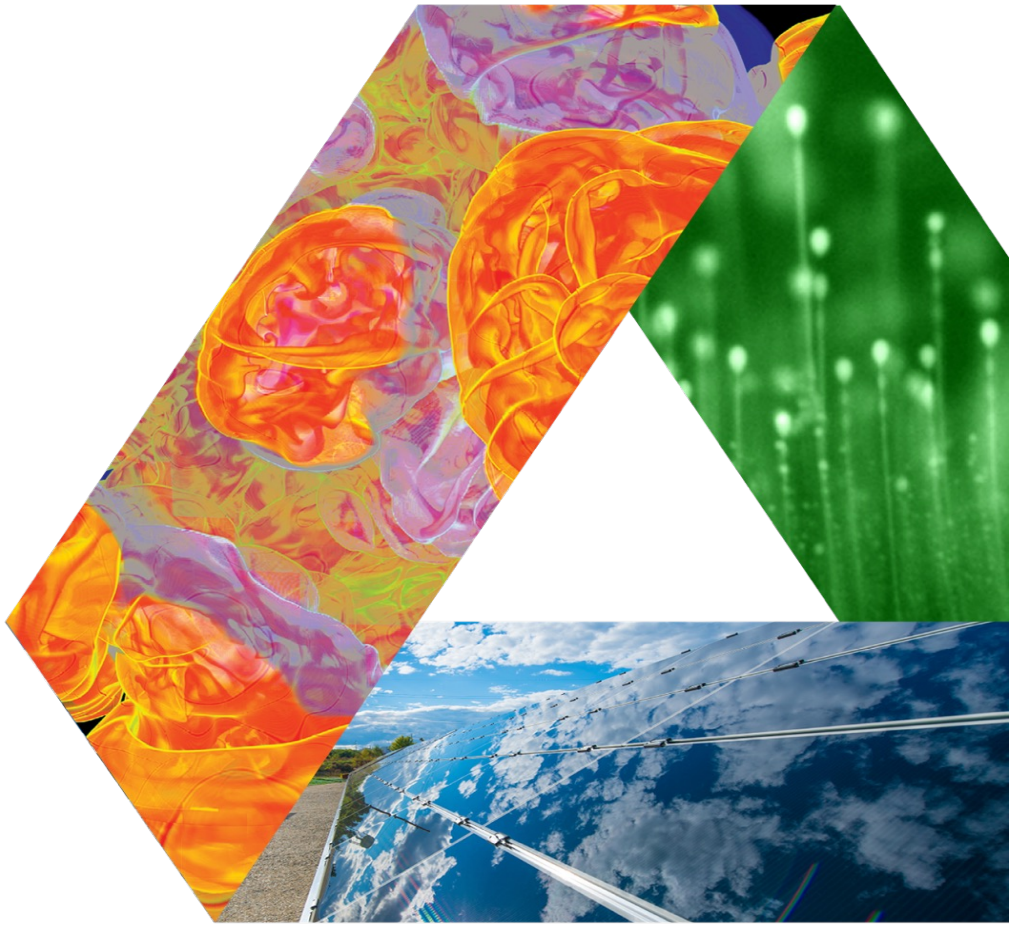
Software Stewardship at ASCR

Workshop on Software Infrastructure for Advanced Nuclear Physics Computing

June 21, 2024

Anshu Dubey
Mathematics and Computer Science Division



Argonne National Laboratory

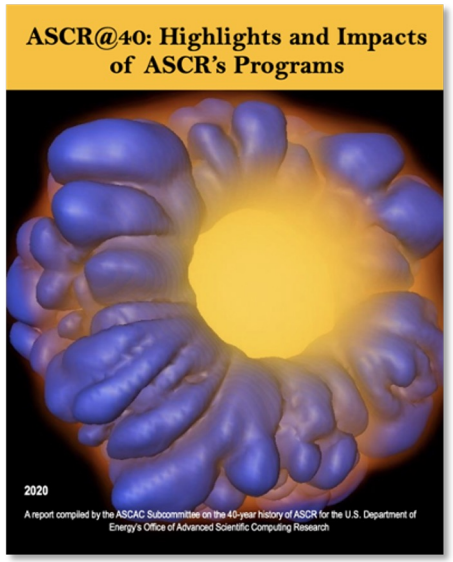
Reusable scientific software provides a foundation for discovery and innovation

ASCR: DOE's Office of Advanced Scientific Computing Research

***ASCR@40: Highlights and Impacts of ASCR's Programs*,** Hendrickson, Messina et al., 2020

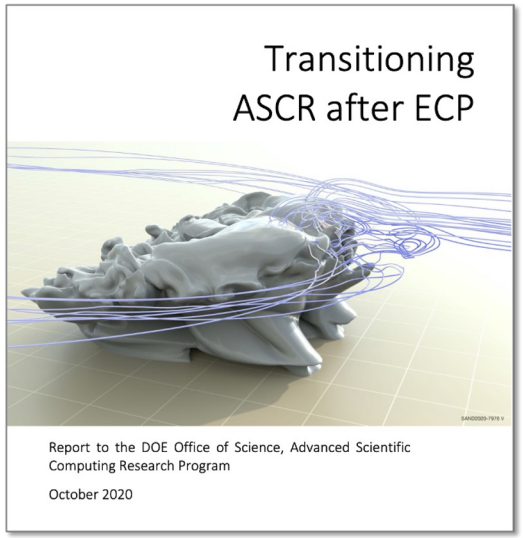
Challenges of the future

- Technology disruptions
- Funding balance
- **Software stewardship**
- Broader partnerships
- Sought-after workforce
- New roles for computing to advance science



***Transitioning ASCR After ECP*,** Giles et al., 2020

- **A. Advancing and Building on ECP**
- **B.** Advancing ASCR Research
- **C.** Current and Future Workforce



Challenge 3: Software Stewardship

The community has long struggled to settle on a good model for sustained support for key elements of the software ecosystem. This issue will get more acute as the Exascale Computing Project winds down and its large, focused software efforts are at risk of being left high and dry with no support for continued development or maintenance. **ASCR needs to recognize that software is really a scientific facility that requires sustained investments in maintenance and support. Improved and sustainable funding mechanisms are required.**

Closely related is the need for investments in improved software engineering practices reflecting profound changes in the way scientific software is developed and maintained. Modern scientific software is increasingly the product of large, dispersed teams and leverages a diverse suite of libraries and tools … Investments in developing, applying, and advancing best practices in software engineering for scientific applications will be essential for continued progress.

Recommendation A1: Create a shared-software stewardship program within ASCR

ASCR should create a comprehensive program that leverages the ECP ecosystem to support and curate shared software.

Stewardship is vital

Our community discussions indicated a strong consensus on the need for some form of software stewardship.

Leadership for the long term

We believe the software hub should be led by DOE laboratory and academic software leaders who possess long-term experience with DOE software development and delivery.

Exascale Computing Project

Maintain **international leadership in HPC**

Promote the **health of the US HPC industry**

Deliver a **sustainable software ecosystem** used and maintained for years to come

Ensure that exascale systems can be used to deliver **mission-critical applications**

ECP EXASCALE COMPUTING PROJECT

7-year, $1.8B

US Department of Energy project funded 1000+ people at national labs, universities, US industries

This research was supported 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.

Application Development

- Develop and enhance the predictive capability of applications, **25 applications, 6 Co-Design Centers**

Software Technology

- Deliver expanded and vertically integrated software stack, **70 unique products**

Hardware and Integration

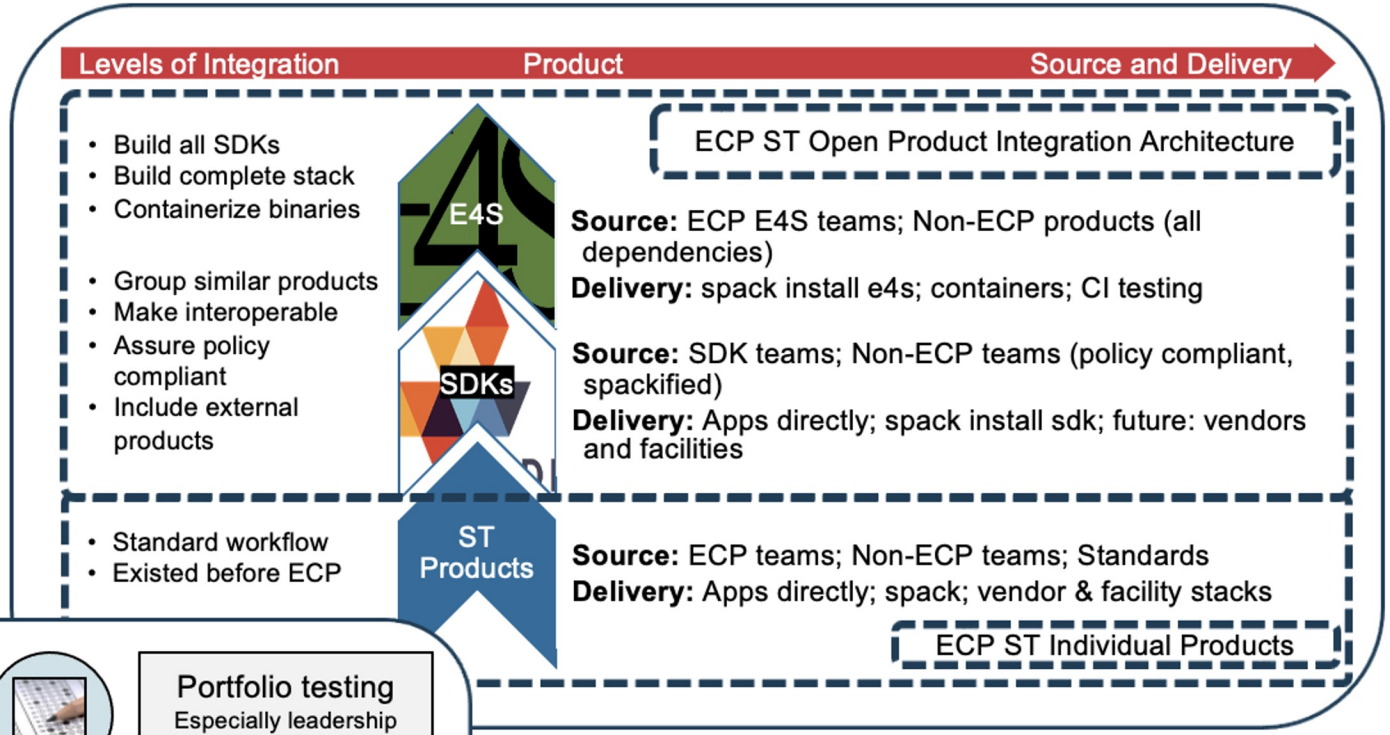
- Application integration and software deployment to facilities, exascale node and system design, **6 US HPC vendors**

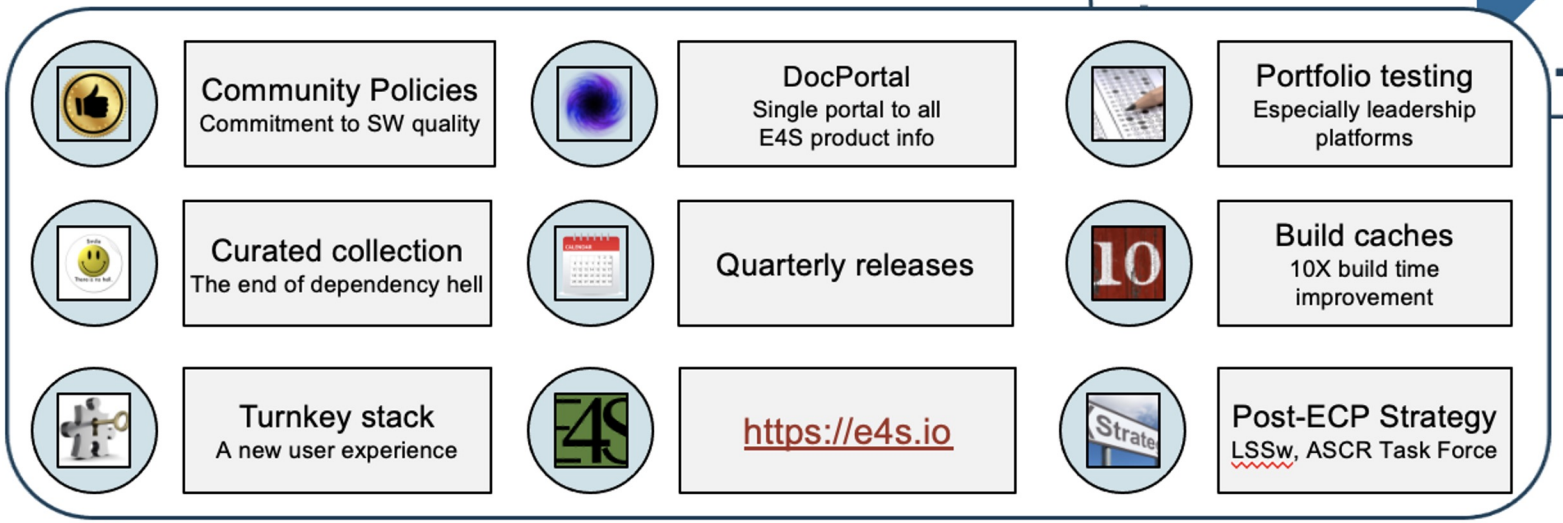
NNSA National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY | Office of Science

A robust HPC scientific software stack – for exascale and beyond

- ECP Software Technology lead: Mike Heroux (SNL)
- E4S: HPC software ecosystem – a curated software portfolio
- A **Spack-based** distribution of software tested for interoperability and portability to multiple architectures
- Available from **source**, **containers**, **cloud, binary caches**
- Not a commercial product – an open resource for all
- Supported by DOE and commercial entities (Paratools)
- Growing functionality: November 2023: E4S 23.11 – 100+ full release products





https://e4s.io

E4S lead: Sameer Shende (U Oregon)

Spack https://spack.io

Spack lead: Todd Gamblin (LLNL)

Stewardship of Next Generation Scientific Software and Technologies (NGSST)
Software Stewardship Organizations (SSOs)

DOE Office of Advanced Scientific Computing Research (ASCR) Post-ECP Projects

COLABS

Training, workforce development, and building the RSE community

CORSA

Partnering with foundations to provide sustainable pathways for scientific software

FASTMATH

Stewardship, advancement, and integration for math and ML/AI packages

PESO

Stewarding, evolving and integrating a cohesive ecosystem for DOE software

RAPIDS

Stewardship, advancement, and integration for data, visualization and ML/AI packages

S4PST

Stewardship, advancement and engagement for programming systems

STEP

Stewardship, advancement of software tools for understanding performance and behavior

SWAS

Stewardship and project support for scientific workflow software and its community

The Consortium for the Advancement of Scientific Software

CASS Basics

- A newly-formed organization
- Sponsored by DOE Office of Advanced Scientific Computing Research (ASCR)
- Established by DOE Software Stewardship Organizations (SSOs)

CASS Goals

- Forum for SSO collaboration and coordination
- Bigger than the sum of its parts
- Vehicle for advancing the scientific software ecosystem

CASS Status

- Defining governance structure
- Establishing community awareness
- Building a team of teams
- Collaborating on outreach

Software Stewardship Organization (SSO) Basics

- Each SSO represents a specific software ecosystem concern
- **Product SSOs:** Programming systems, performance tools, math packages, data/viz packages
- **Portfolio SSO:** Curating & delivering software stack to the community
- **Community SSOs:** Workforce, partnerships

Engage with CASS

- Participate in June 11-13 CASS Community BOF Days: https://cass.community/bofs
- Visit https://cass.community

COLABS: Collaboration for Better Software (for Science)

Motivations and Objectives

Helping software teams produce better software

- Training and complementary resources on software productivity and best practices
- Coordination of training development and delivery for CASS
- Advocacy for better software practices with sponsors, project leaders, and developers

Helping grow and enrich the workforce of software professionals

- Building the research software engineering (RSE) community of practice (CoP) in the national laboratories
- "Onboarding training" to support broader candidate pools for software professional positions
- Filling the pipeline: outreach to colleges for internships and careers

Training and Related Resources

An extensive training program for software productivity and best practices

- Building on the successful Better Scientific Software (BSSw) tutorial series, originated in the IDEAS project (> 30 events since 2016)
- Collaborating with related training efforts in the community

Providing related resources

- HPC Best Practices webinar series (monthly cadence)
- Better Scientific Software (BSSw.io) resource site (w/ PESO)

Coordination of training for CASS

Advocacy for better software practices

Building the DOE RSE CoP

Build interactions, knowledge sharing among RSEs at national labs

- Encouraging social interactions
- Sharing of technical and professional knowledge and experiences
- In collaboration with U.S. Research Software Engineer Association (US-RSE)

Other Workforce-Building Efforts

Onboarding training

- Expand applicant pools by developing training curricula to allow new hires to acquire needed skills
- Both technical and "soft" skills

Retention-enhancing practices for software projects

- Identify and promote practices software projects and organizations can implement to improve employee satisfaction and retention

Filling the pipeline

- Outreach to colleges (esp. MSIs) to promote internships and careers in research software engineering

Team

Argonne National Laboratory: Anshu Dubey (Lead PI), Rinku Gupta
Oak Ridge National Laboratory: David Bernholdt (PI), Greg Watson
Lawrence Berkeley National Laboratory: Dan Gunter (PI), Keith Beattie
ASCR PM: David Rabson

Note: Phase-2 proposal is pending

Argonne NATIONAL LABORATORY
OAK RIDGE National Laboratory
BERKELEY LAB

CORSA: Center for Open-Source Research Software Advancement

CORSA is a five-year post-ECP software-ecosystem stewardship and advancement project. In partnership with the Consortium for the Advancement of Scientific Software (CASS), CORSA will become a community of practice that creates pathways for open-source scientific software projects to avail themselves of resources for long-term growth, stewardship, advancement, and innovation.

Goals

Create pathways to open source software foundations

Empower communities with software sustainability metrics

Facilitate cross-cutting activities to address the needs of unique and diverse communities

Provide objective guidance to software communities

Activities

- **Foundation Forum**: We lead CASS efforts for enabling software projects to leverage open source software foundations as a pathway toward sustainability
- **Software Sustainability Metrics**: We will work with the community to develop software sustainability metrics and coordinate mechanisms to collect and utilize metrics to best serve the needs of CASS members
- **Best Practices**: We gather successful project and community experiences to create a repository of best practices, example documents, and guidance that can be used by software projects in their sustainability efforts
- **Guidance and Training Resources:** Curate and create guidance and training resources to share and disseminate strategies for sustainability activities

Software Stewardship in the RAPIDS and FASTMath Institutes

Objective

The objectives of stewardship activities within RAPIDS and FASTMath are to advance and steward a selected set of production software products for use in leadership computing providing key capabilities to DOE science teams, to help bring relevant new software products to a production state, and to foster the integration of software tools into effective solutions that help achieve DOE mission objectives.

Significance and Impact

A robust software ecosystem is a critical enabler of DOE science, helping adapt to the evolution of the underlying hardware technology ecosystem and evolving DOE mission needs.

Technical Approach

- Utilize stakeholder input to identify priority software products, working with other SSOs to identify key capability gaps and avoid duplication of effort
- Fund product-specific activities that improve the software in terms of impact, quality, and/or sustainability
- Leverage the wealth of knowledge in the ASCR and technical computing communities to identify promising potential products for “incubation”
- Actively collaborate with our peer SSOs to train software teams on best practice, to integrate into the broader software ecosystem, to gain access to CI/CD resources, and to build connections between teams
- Promote equity and inclusion as an intrinsic element to advancing scientific excellence

Name	Summary
Data	
ADIOS	ADIOS provides a simple, flexible way for scientists to describe the data in their code that may need to be written, read, or processed.
DIY	Block-parallel library for writing scalable parallel algorithms
HDF5	Parallel I/O library for high performance access to self-describing HDF5 datasets
Paraview/Catalyst	ParaView is a tool for post hoc visualization at scale. Catalyst is the associated in situ library.
PnetCDF	Parallel I/O library for high performance access to self-describing netCDF datasets
VisIt/Ascent	VisIt is an open source, interactive, scalable, visualization, animation and analysis tool. Ascent is a lightweight in situ library leveraging VTK-m for GPU support.
VTK-m	VTK-m is a performance portable visualization library leveraged by production tools such as ParaView and VisIt.
zfp	Open-source library for compressed floating-point and integer arrays that support high throughput read and write random access.
Applied Math	
AMReX	Framework for massively parallel, block-structured adaptive mesh refinement applications
Ginkgo	High-performance linear algebra library for manycore systems, with a focus on solution of sparse linear systems
MAGMA	Matrix Algebra on GPU and Multi-core Architectures (MAGMA) is a collection of next-generation linear algebra libraries for heterogeneous computing
hypre	Linear solvers library with emphasis on multigrid for structured and unstructured problems
Kokkos Kernels	On-node, portable, sparse/dense linear algebra library based on Kokkos programming model.
MFEM	Light-weight, scalable building blocks for implemeting finite element algorithms
PETSc	Suite of data structures, linear solvers, preconditioners, non-linear solvers for distributed memory scientific simulations
STRUMPACK	Solvers for dense rank-structured or sparse linear systems with support for variety of rank-structured formats
SUNDIALS	Nonlinear solvers and time integration library
SuperLU	Sparse direct solver library providing a robust solution approach for hard to solve sparse linear systems.
Trilinos	Framework with linear solvers, preconditioners, non-linear solvers, partitioning tools, and discretization libraries to support science simulations
libCEED	libCEED provides fast algebra for element-based discretizations, designed for performance portability, run-time flexibility, and clean embedding
AI/ML	
DeepHyper	Scalable neural architecture and hyperparameter optimization for deep neural networks on leadership-class machines
Tools	
ERT (Roofline)	Automates generation of roofline model for architecture-specific performance analysis and optimization
TAU	Portable profiling and tracing toolkit for performance analysis of parallel programs written in Fortran, C, C++, UPC, Java, Python.

FY24 Institute supported software stewardship recipients

PESO: Partnering for Scientific Software Ecosystem Stewardship Opportunities

About PESO

- PESO is a five-year post-ECP software-ecosystem stewardship and advancement project. In partnership with the Consortium for the Advancement of Scientific Software (CASS), PESO will establish and steward a sustainable scientific software ecosystem comprising libraries and tools that deliver the latest high-performance algorithms and capabilities for DOE mission-critical applications.

Key PESO goals

- PESO will enable applications to realize 100X improvement in both high-end capabilities and energy efficiency by leveraging accelerator devices
- PESO will emphasize software product quality, the continued fostering of software product communities, and the delivery of products, working with CASS

Key PESO Activities

- **Partnerships**: We lead CASS efforts for diverse and inclusive workforce with sustainable career paths. We shepherd BSSw Fellows Program and BSSw.io portal.
- **Services**: We provide services including software product management, integration, and delivery, as well as software quality assurance and security.
- **Products**: We deliver & support products via Spack & E4S, provide porting & testing platforms leveraged across product teams to ensure code stability & portability.

PI: Michael Heroux, Sandia National Laboratories
Collaborating Institutions: ANL, Berkeley, BNL, Kitware, LLNL, LANL, ORNL, PNNL, SNL, SHI, UO
ASCR Program: Software Stewardship and Advancement
ASCR PM: William Spotz
Resources: https://pesoproject.org, https://e4s.io, https://hpsf.io

The PESO Project exists to preserve, sustain, and advance the investments made by the Exascale Computing Project in a robust, versatile, and portable HPC software ecosystem and the people who make the ecosystem effective.

PESO is unique in its organization by composing itself of many members of other software stewardship organizations (SSOs) to best ensure tight integration across the SSOs and support the mission of the Consortium for the Advancement of Scientific Software (CASS).

S4PST – Stewardship for Programming Systems and Tools

Objective: Advancement of programming systems for the next generation high performance computing systems and its seamless integration with emerging AI technologies for science.

Description: Two-fold strategy

1. Preserve the legacy of the US Department of Energy Exascale Computing Project (ECP) critical HPC programming systems
2. Incubate the integration of HPC with the latest industry-driven AI technologies revolutionizing the broader computing landscape (e.g. large language models, LLMs).

Approach

Programming Systems are the first point of contact in human-computer interactions, and never ceased to evolve along with hardware and software technology

- Steward the legacy and strategic evolution of ECP critical programming systems
- Act upon technical, economical, and social requirements to integrate HPC and AI for science
- Provide flexibility in defining scalable activities to maximize the impact of our portfolio on ASCR
- Balance value and growth activities through coordination with ASCR stakeholders and those in the broader computing landscape.
- Provide an aggressive strategy to diversify the highly-specialized pipeline of programming systems stewards to promote inclusive and equitable research.

S4PST Product Portfolio

Product	*Product Description*	*Product Lead*	*Lead Institution*
OpenMP/OpenACC	*Portable Parallel Programming Extension and runtime for C, C++ and Fortran*	*Sunita Chandrasekaran*	*UD*
LLVM	*Focus on HPC support for the de facto standard Open-Source Compiler*	*Johannes Doerfert*	*LLNL*
Kokkos/C++	*Performance Portable Programming Systems for modern C++*	*Christian Trott*	*SNL*
Fortran	*Focus on new node parallel computing features*	*Damian Rouson*	*LBNL*
OpenMPI	*MPI implementation contributed by several institutions and vendors*	*Thomas Naughton III*	*ORNL*
MPICH	*First Open Source MPI implementation. Basis for several vendors' MPI*	*Yanfei Guo*	*ANL*
Legion	*Distributed Asynchronous Task-Parallel Runtime*	*Pat McCormick Alex Aiken*	*LANL SLAC*
UPC++ & GASNet-EX	*Alternative distributed programming model and runtime*	*Paul Hargrove*	*LBNL*
HIP	*AMD's portable accelerator programming runtime*	*Brice Videau*	*ALCF*
SYCL	*Portable modern C++ programming systems with compiler support.*	*Thomas Applencourt*	*ALCF*

STEP Overview – Software Tools for Understanding Performance and Behavior

Software Tools Ecosystem Project (STEP) – https://ascr-step.org

Scientific Achievement

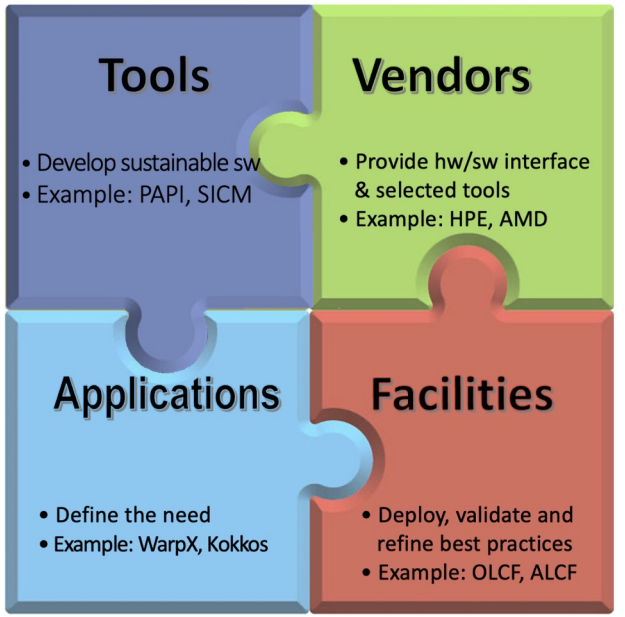
The Software Tools Ecosystem Project (STEP), a member of CASS, enables scientists, facility operators, and vendors to unlock new scientific discoveries by providing the tools necessary to understand and enhance the behavior of scientific applications at scale.

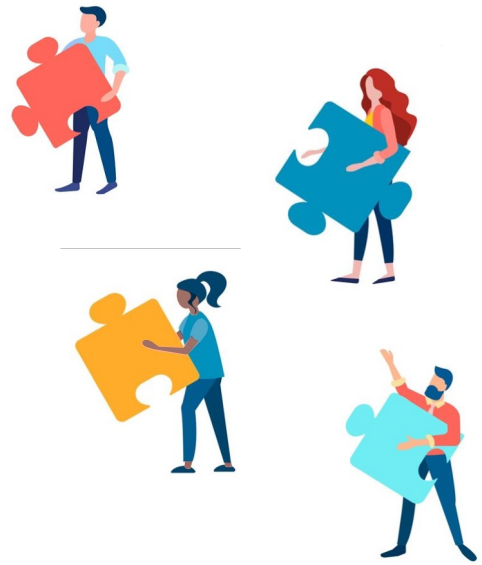
Significance and Impact

STEP provides stewardship and advancement for critical tools and supporting software that perform monitoring, analysis, and diagnosis of performance and behavior of codes. Examples include application profilers, tracing tools, system monitors, etc. As computers have increased in complexity and scale, using them effectively has become much more difficult. STEP addresses this acute need through sophisticated and openly available software.

Initial Funded Software Packages

Software Tool	PI
HPCToolkit	John Mellor-Crummey, Rice Univ.
PAPI	Heike Jagode, Univ. of Tennessee
Dyninst	Barton Miller, Univ. of Wisconsin
TAU	Sameer Shende, Univ. of Oregon
Darshan	Shane Snyder, Argonne Nat. Laboratory





Cutting edge software tools are closely bound to architectures, system software, and applications in ways that other types of software are not. STEP relies on proactive co-design of interoperable tools and deep interaction with diverse stakeholders to address this challenge. This is accomplished via recurring community meetings and input from vendor representatives, application scientists, and facility operators on the STEP Oversight Council.

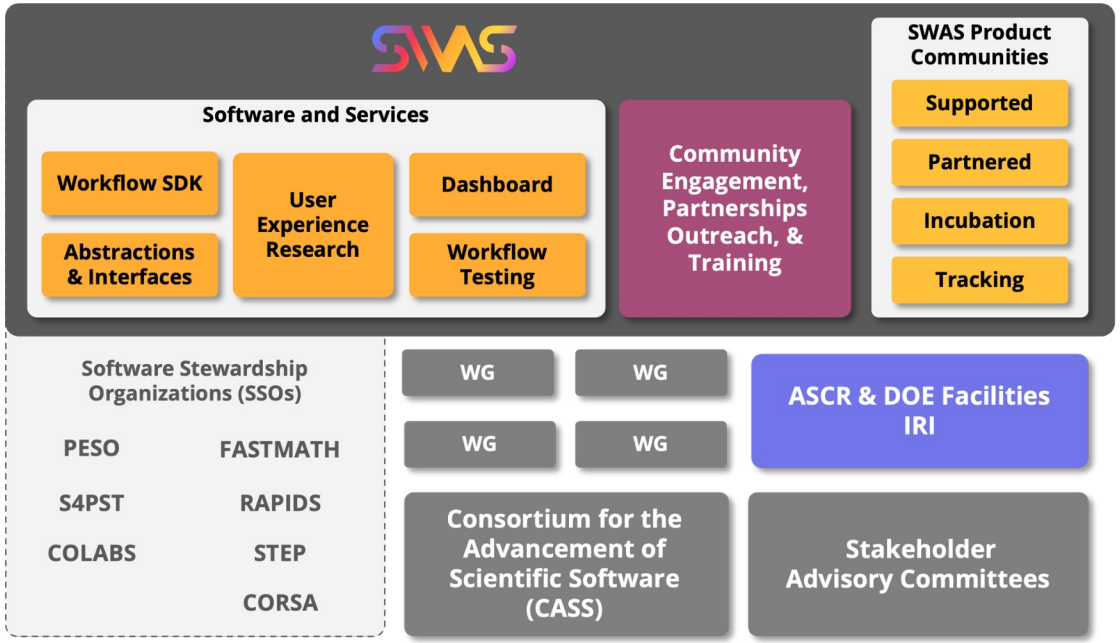
STEP Consortium Leads: Terry Jones, ORNL (Director), Philip Carns, ANL (Deputy Director)
ASCR Program: Next Generation Scientific Software Technologies (NGSST)
ASCR PM: David Rabson
STEP website: https://ascr-step.org/

STEP SOFTWARE TOOLS ECOSYSTEM PROJECT

SWAS: Stewardship and Advancement of Workflows and Application Services

Objective

SWAS is a pivotal initiative addressing the complex **workflow** needs of the **DOE science community**. It aims to centralize activities around workflow software, including *research*, *development*, *education*, and *training*. By focusing on simulations, real-time data analysis, and AI/ML integration, SWAS seeks to overcome the fragmentation and inefficiencies in the current scientific workflow ecosystem. Its mission is to steward, support, and advance critical **workflow software and services** across all DOE science disciplines.



Significance and Impact

SWAS targets enhancing core capabilities widely required by the community and aims to foster robust community engagement, partnerships, outreach, and training initiatives. SWAS is necessary to bringing the IRI's strategic vision to fruition by supporting the essential software infrastructure and services needed to integrate computational and experimental research facilities seamlessly. SWAS is a vital bridge between ASCR computing facilities and research and software development efforts, enhancing the DOE's capacity to tackle pressing scientific challenges through improved computational efficiency and innovation.

Technical Approach

- ***Community Engagement and Workforce Development:*** SWAS prioritizes building a vibrant, engaged community. It focuses on workforce support, disseminating best practices, and training initiatives to cater to the diverse research needs across DOE science disciplines.

- ***Software Evaluation and Lifecycle Management***: SWAS will employ workflow-specific metrics for evaluating software, ensuring its relevance and efficacy in addressing the dynamic requirements of scientific research. Our active participation in the governance of the S3C ecosystem is pivotal to not only facilitate the continuous adaptation and improvement of the workflow ecosystem but also underscores our dedication to supporting the DOE's mission to lead in scientific innovation.

- ***Support for Critical Workflow Software:*** By identifying and supporting software that offers core capabilities essential to the scientific community, SWAS addresses the diverse needs of scientific workflows, from simulation orchestration to data analysis and AI/ML.

Launched of the High Performance Software Foundation (HPSF) at ISC

HIGH PERFORMANCE SOFTWARE FOUNDATION

Scientific Achievements

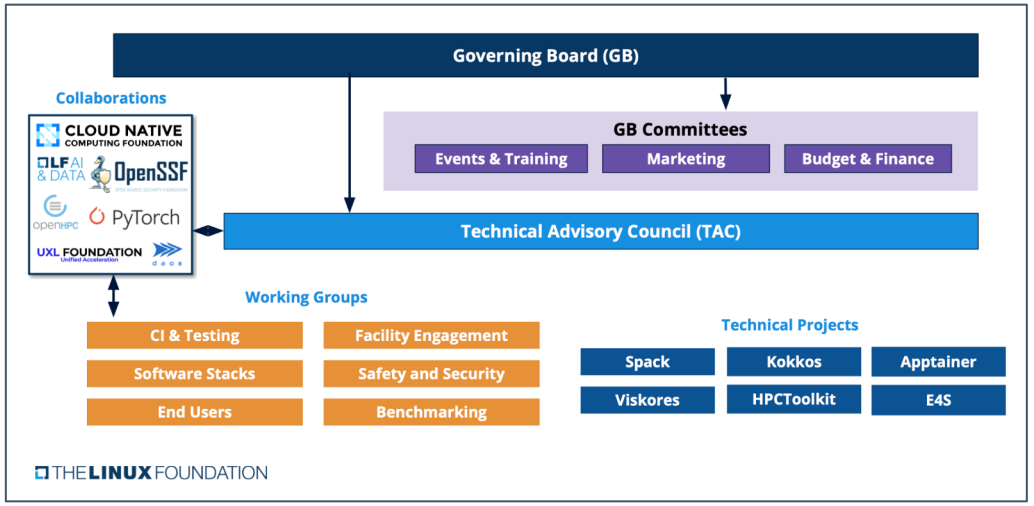
- NNSA led the formation of HPSF over the past 18 months, in close collaboration with DOE/ASCR, industry, and the Linux Foundation
- 15 founding members, 6 initial projects from industry, academia, labs around the world
- HPSF provides open source projects with:
  - A neutral home
  - Collaborations
  - Open governance
  - Funds for project infrastructure, working groups, events, other initiatives

Significance and Impact

- Initial membership raised more funds than expected: **~$1M annual budget**
- ISC kickoff BOF session was standing-room only; generating much excitement
- Expect to grow membership and projects over time
  - 2 NNSA, 2 ASCR, 2 European projects in the pipeline
  - Expecting at least 2 more general members within the year

Approach

- Separate financial (GB), technical (TAC), and project governance
- Interact directly with key projects in the HPC ecosystem
- Grow contributor base through increased adoption, training, events, outreach
- Build a portable, accelerated software stack for HPC and beyond



Premier Members

aws, Hewlett Packard Enterprise, Sandia National Laboratories, Lawrence Livermore National Laboratory

General Members

AMD, intel, NVIDIA, Argonne NATIONAL LABORATORY, OAK RIDGE National Laboratory, Los Alamos NATIONAL LABORATORY, kitware, cea

Associate Members

UNIVERSITY OF OREGON, UNIVERSITY OF MARYLAND, CDAC

PI(s)/Facility Lead(s): Todd Gamblin, Christian Trott
Collaborating Institutions: NNSA/ASC, PESO, CORSA, Linux Foundation
DOE Programs: NNSA/ASC, ASCR/NGSST
DOE PMs: Si Hammond, William Spotz, David Rabson, Hal Finkel
Resources: https://hpsf.io

National Nuclear Security Administration

The International Post-exascale Project

Workshop Topics

- ❑ AI for Scientific Computing
- ❑ Software production and management
- ❑ Co-design, benchmarks, mini-apps, proxy and evaluation
- ❑ Digital continuum and data management

Challenges and Opportunities

Challenges

- Training data
  - Curating datasets
  - Datasets need their own readers
- Complex workflows
  - Traditional HPC simulations getting more complex
  - Scientific workflows have stages and interoperability challenges
- Tools have gaps

Opportunities

- Defining ways of ingesting data
  - Think of datasets also as legacy to be handled
  - Learn from those who already make their data publicly available
  - Catalogs with readers and documentation
- Common framework for integrating stages of workflows
  - Something like a pytorch for scientific computing

An Example of Benefits of Building a Framework

	astro-physics	cosmo-logy	CFD/ FSI	HEDP	solar physics	recon-nection	star fo-rmation	combus-tion
compress-ible hydro	1998	*		*	*			*
burn	1999							*
MHD	2002	*		*	*	*	*	
elliptic solver	*	2001 2001	*				*	
particles	*	2002	*	*		*	*	*
bittree	*	*	2012	*				
HYPRE interface			*	2011				
radiation	*	*		2011				

Dubey A, Tzeferacos P, Lamb DQ. The dividends of investing in computational software design: A case study. *The International Journal of High Performance Computing Applications*. 2019;33(2):322-331. doi:10.1177/1094342017747692

An Example of Benefits of Building a Framework

	astro-physics	cosmo-logy	CFD/ FSI	HEDP	solar physics	recon-nection	star fo-rmation	combus-tion
compress-ible hydro	1998	*		*	*			*
burn	1999							
MHD	2002	*						
elliptic solver	*	200 200						
particles	*	200						
bittree	*	*						
HYPRE interface								
radiation	*	*						

Domain	initial effort (person years)	total effort (person years)	release date (year)	first publication (year)	total publications
Supernova,Astro	16	60	2000	2000	427
Cosmology	2.5	20	2001	2001	257
Solar physics	0.5	5	2000	2002	70
HEDP	8	28	2010	2010	38
Core-collapse	1	5	2012	2008	20
CFD/FSI	3	5		2012	5

Dubey A, Tzeferacos P, Lamb DQ. The dividends of investing in computational software design: A case study. *The International Journal of High Performance Computing Applications*. 2019;33(2):322-331. doi:10.1177/1094342017747692

Impact of Well-designed Codes on Science