COLABS

Collaboration for Better Software (for Science)

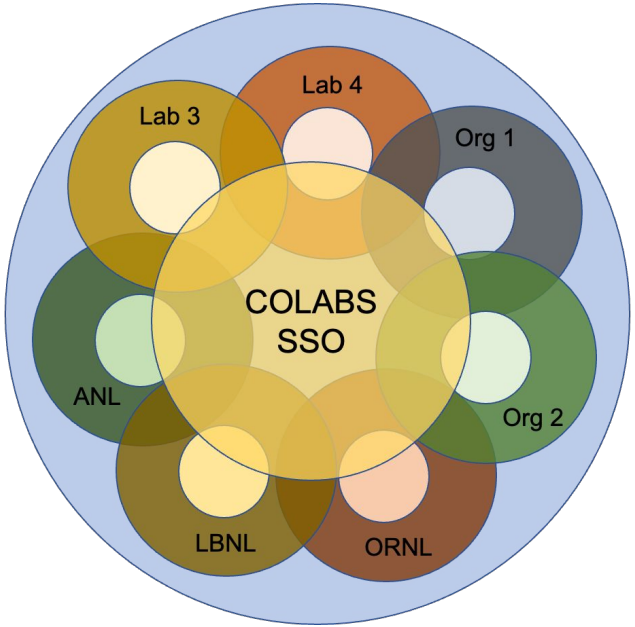
Anshu Dubey, David Bernholdt, Dan Gunter
Kevin Harms, Bronson Messer, Richard Gerber, John MacAuley

https://colabs-science.github.io

Our Vision

The raison d'etre for COLABS is the stewardship of scientific software.

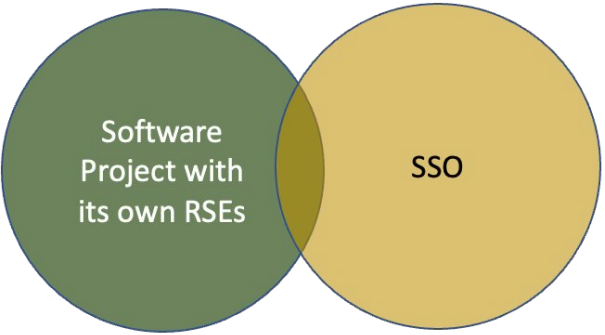
- Research software engineers (RSEs) are central to our approach to the stewardship of scientific software
- We are about building the community of people who are capable and available to support the stewardship of our scientific software, including RSEs, though training, mentoring, advocacy, etc.
  - COLABS is *not* about establishing RSE organizations at specific institutions
- The COLABS model is scalable to easily include additional institutions, in support of additional sponsors



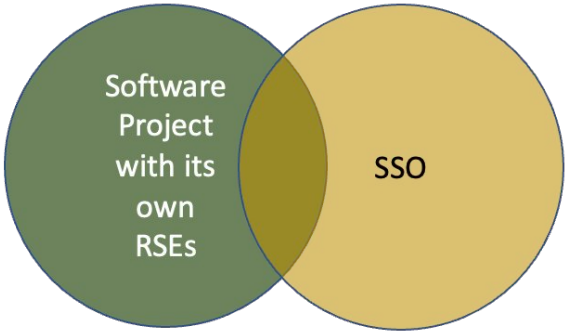
Flexibility to Support the Needs of Client Projects

- Flexibility to address each project's needs on an individual basis
  - Network of people with experience in software development and stewardship
  - Guide client projects in identifying and prioritizing their needs
- Essential level of effort that any client project can expect
- Ability to request additional resources for higher effort/limited duration stewardship activities to address specific needs
- Model for embedded RSEs

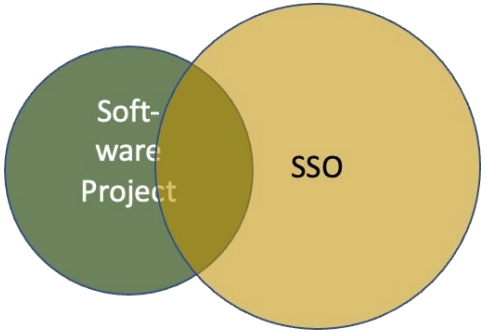
Models of Engagement



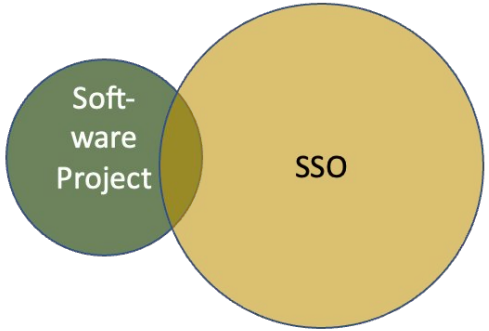
Minimal engagement through entrenched RSEs



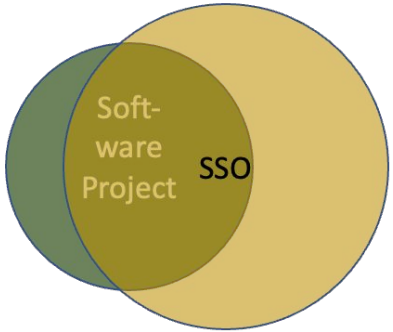
Use of standard maintenance from SSO



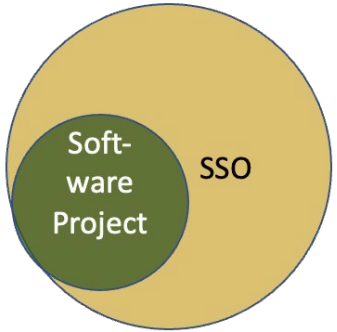
Use of advanced RSE services from the SSO



SSO acts as incubator, training the project personnel



Project gets research funding for adding new features, SSO maintains



SSO assumes full maintenance of an otherwise orphaned project

sustaining **workflows**
& application **services**

https://swas.center

OAK RIDGE National Laboratory | Argonne NATIONAL LABORATORY | Brookhaven National Laboratory | BERKELEY LAB | Lawrence Livermore National Laboratory | U.S. DEPARTMENT OF ENERGY | Office of Science

Overview

SWAS brings together **academia**, **national labs**, and **industry** to create a sustainable software ecosystem supporting the myriad software and services used in **workflows** as well as the workflow orchestration software itself

SWAS will grow, support, and sustain the ecosystem spanning the full range of **analysis**, **simulation**, **experiment**, and **machine learning workflows**

SWAS will ensure that researchers can rely on robust, portable, scalable, secure, and interoperable workflows software and application services

Objectives

> "Identify critical software and develop a plan for sustaining this software that is tailored to the needs of this unique software ecosystem"

Community-endorsed sustainability model

SWAS will advance and sustain workflows and application services development, with entrusted **validation** and **verification** capabilities

Target Software

Workflow Systems

Data Management Frameworks

Visualization Frameworks

AI/ML Tools *(used in modern workflows)*

Application services *(including Cloud services)*

We emphasize that "workflows" in this context represents the ***broad set of software and services*** *users need to configure, orchestrate, and operate modern analysis, modeling, and simulation campaigns*

Stakeholder Communities

sustaining **workflows** & application **services**

Workflows and application services

focus on general and specific domains, non-expert and expert users, and offer configuration-based interfaces, graphical interfaces, domain-specific languages, or programming language libraries or APIs

Science and engineering communities

understand their current, imminent, and future workflow needs and challenges, and provide guidance for application, infrastructure, and software development

Computing centers and facilities operators

support the use and deployment of workflow and application services, and provide training to foster proper adoption of workflow tools and therefore offering pathways for sustainability

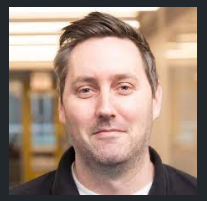
Get Involved

The SWAS seedling effort is planning to organize an in-person workshop during **Summer 2023**

contact@swas.center

Leadership Team

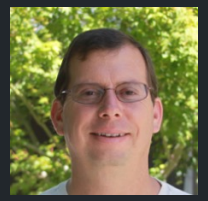
Rafael Ferreira da Silva (ORNL)



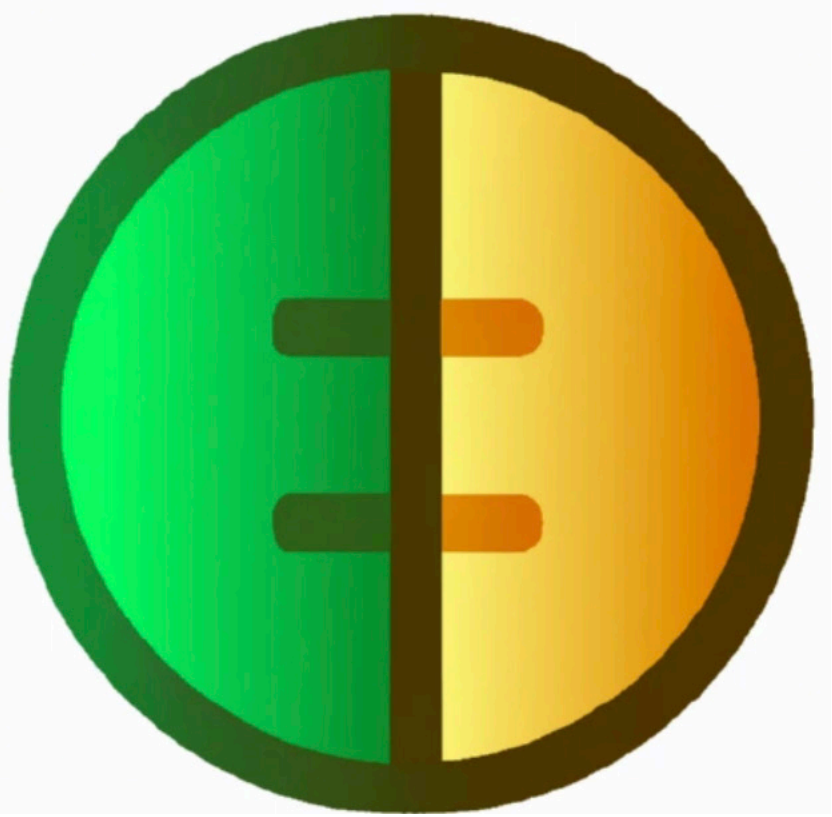
Kyle Chard (ANL)

Lavanya Ramakrishnan (LBL)

Shantenu Jha (BNL)



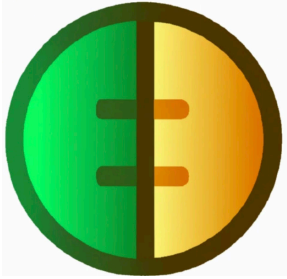
Dan Laney (LLNL)



Toward a Post-ECP Software Sustainability Organization (PESO)

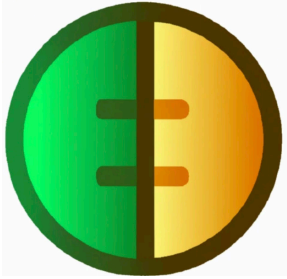
- Michael Heroux (Sandia National Laboratories; PI)
- James Ahrens (Los Alamos National Laboratory)
- Todd Gamblin (Lawrence Livermore National Laboratory)
- Timothy Germann (Los Alamos National Laboratory)
- Xiaoye Sherry Li (Lawrence Berkeley National Laboratory)
- Lois Curfman McInnes (Argonne National Laboratory)
- Kathryn Mohror (Lawrence Livermore National Laboratory)
- Todd Munson (Argonne National Laboratory)
- Sameer Shende (University of Oregon)
- Rajeev Thakur (Argonne National Laboratory)
- Jeffrey Vetter (Oak Ridge National Laboratory)
- James Willenbring (Sandia National Laboratories)

Funding for Product Teams



- ASCR (and ASC) program managers will fund product development teams based a decision-making process they own
- Product teams: Receive funds based on sponsor process
- Seed teams: Focused on organizational approaches to enhance ecosystem sustainment
- Bottom line:
  - Our sponsors will decide product funding (and know that ECP ends soon)
  - Seed projects will develop plans to work with product teams, sponsors, industry, and others to make ecosystem sustainment real

PESO Goals Sketch

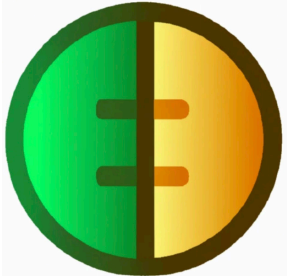


- Collaboratively steward, facilitate and aggregate activities, processes, resources, relationships, and more
  - Across DOE-sponsored teams, and teams of teams
  - Engaging sponsors, facilities, industry, and community organizations
- Represent the collective interests of the DOE software community
  - What this means depends on the planning process we are in right now
  - Goal is to represent interests that are cross-cutting, not addressed elsewhere
- Provide large-scale infrastructure
  - Software portfolio management at E4S level
  - Spack integration, CI testing, containers, other software ecosystem needs
  - Portfolio lifecycle management

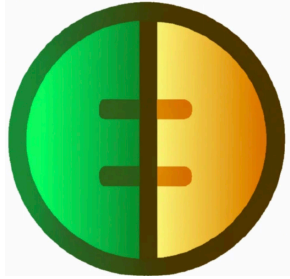
PESO Governance Principles

- Place decision-making ownership at the lowest level possible
- Integrate across lab, industry, and university collaborators
- Distribute cost and benefit sharing
- Create a diverse and inclusive workforce with sustainable career paths
- Commit to productivity and sustainability improvement
- Engage with external community members

How to Collaborate with PESO

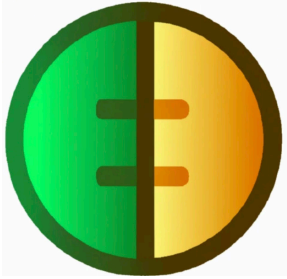


- Software Product Communities (aka, SDKs)
  - PESO anticipates engaging with teams of product teams
  - We anticipate SPCs will self-organize and have community-specific governance
  - We anticipate SPCs to include DOE-sponsored and commercial/community software
- Communities of Practice
  - PESO anticipates engaging with community leaders in important cross-cutting efforts
  - Examples include:
    - RSEs: (e.g., IDEAS, HPC Best Practices webinars)
    - Community outreach (e.g., Center for Scientific Collaboration and Community Engagement (CSCCE)
    - Software foundations (e.g., NumFOCUS, Linux Foundation)
    - Workforce development (e.g., US RSE, BSSw Fellows, and Sustainable Research Pathways)



Sponsors	How are software projects selected for sustainability funding? How is sustainability success defined and measured? When and how is software support transitioned to industry?	PESO Planning Questions	How do we identify and prioritize lib/tool features? How do we co-design/develop features & APIs? How do we deliver emerging and stable capabilities?	Application development communities
Math/CS R&D projects	When is an R&D effort ready for transition to sustainability? How can R&D efforts improve productivity & sustainability? How does an R&D effort transition to a sustainability focus?		How do we co-design/develop new features? How do we deliver capabilities to commercial partners? How do we determine support transition from DOE?	Commercial software vendors
DOE computing facilities	How are facilities requirements captured and prioritized? How can facilities provide input for software project selection? How can software products be best provided to facilities users?		How do we access and manage NDA information? How do we co-design/develop features & APIs? How do we deliver capabilities to vendor partners?	Computer system vendors
Third-party integrators	How can we make DOE products easier to adopt outside DOE? How do we manage delivery of DOE software to third-parties? How do we support third-parties in the use of DOE software?		How do we optimize software ecosystem value? How do we improve productivity & sustainability? How do we support software distribution?	Library & tool development teams
Broader community	How do we communicate beyond direct collaborators? How do we account for broader requirements & initiatives? How do we train and support community users?		How can we cultivate a diverse & inclusive workforce? How can we attract, retain and invest in staff? How do we establish sustainable career paths?	Workforce development

PESO Planning: 3 Complementary Opportunities



- Provide input via the PESO Planning Input Google Form
  - https://bit.ly/peso-2023-input
  - We strongly encourage you to provide your input by May 15 for inclusion in the June 8 – 9 workshop
- Engage in PESO Community Discussions (https://lssw.io/PESO)
  - PESO will participate in LSSw Townhalls (see https://lsswi.o)
- Attend the PESO Community Workshop
  - https://bit.ly/peso-workshop-june2023
  - Anyone from the community is welcome as space is available

OAK RIDGE National Laboratory

Introducing STEP – Sustainable Tools Ecosystem Project

Terry Jones, PI

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

April 20, 2023

U.S. DEPARTMENT OF ENERGY

ASCR FOA #2844 Informational Meeting

Sustainable Tools Ecosystem Project (STEP) Team

ROLE	PERSPECTIVE	NAME	ORGANIZATION
Lead-PI	Tools	Terry Jones	Oak Ridge National Lab
Co-Investigator	Tools	James Brandt	Sandia
Co-PI	Tools	Philip Carns	Argonne National Lab
Co-PI	Vendors	James Custer	Hewlett Packard Enterprise
Co-PI	Vendors	Kshitij Doshi	Intel Corporation
Co-PI	Tools	Ann Gentile	Sandia
Co-Investigator	Facilities	Kevin Harms	Argonne National Lab
Co-PI	Tools	Heike Jagode	University of Tennessee
Co-PI	Tools	Mike Jantz	University of Tennessee
Co-PI	Tools	Matthew Legendre	Lawrence Livermore Natl Lab
Co-PI	Vendors	Keith Lowery	Advanced Micro Devices
Co-PI	Tools	John Mellor-Crummey	Rice University
Co-PI	Tools	Barton Miller	University of Wisconsin
Co-PI	Vendors	José Moreira	IBM
Co-PI	Applications	Erdal Mutlu	Pacific Northwest Natl Lab
Co-Investigator	Facilities	Phil Roth	Oak Ridge National Lab
Co-PI	Tools	Sameer Shende	University Oregon
Co-Investigator	Tools	Shane Snyder	Argonne National Lab
Co-PI	Tools	Galen Shipman	Los Alamos National Laboratory
Co-PI	Tools	Devesh Tiwari	Northeastern University
Co-PI	Applications	Theresa Windus	Ames National Lab

STEP Targets A Specific Segment of Scientific Computing

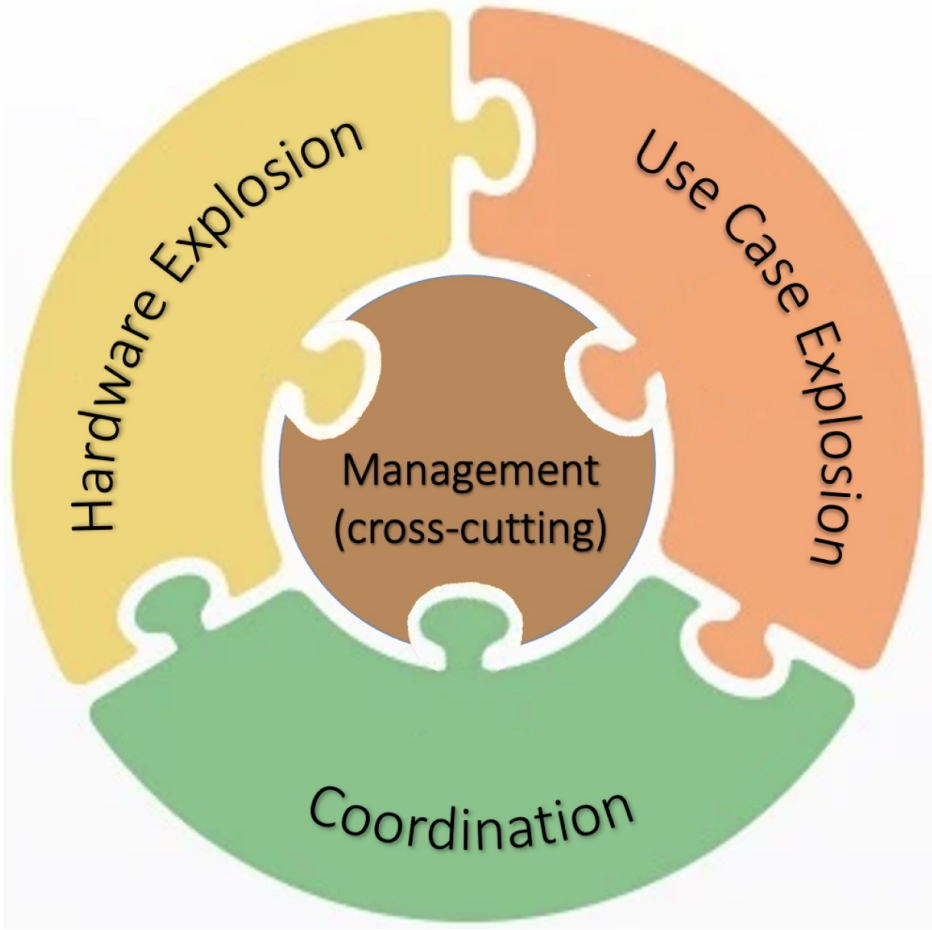
- WHAT ARE *TOOLS*?
  - We define Tools to mean "*the collection of software that can be used to both understand performance bottlenecks and optimize performance and resource efficiency.*"
- WHAT'S SO IMPORTANT ABOUT TOOLS?
  - *As computers have increased in complexity and scale, using them effectively has become much more difficult.*
  - In addition to their role in enabling supercomputer performance (a decisive determinant of scientific discovery), these tools provide essential feedback to users, operations staff, and system and application software developers.

How Are Tools Unique?

- Tools are closely bound to architectures and system software in ways that other types of software, such as libraries and scientific applications, are not.
- For example, a tool that tracks how an application uses computing resources must be able to measure low-level architectural events and metrics and relate them to program progress and source code.
- The need for tools is most acute for understanding code performance on systems that push the boundaries of technology and scale, but these systems' novelty makes them extremely difficult for tool developers to support when first deployed.

Can Tools Transition from Reactive to Proactive?

Four Key Challenges Faced By STEP



STEP – Strategy

Building An Enduring Ecosystem for Vital Tools

- "STEP will bring together a diverse community of High Performance Computing (HPC) tools developers and stakeholders to **develop plans for the sustainability of the HPC tools ecosystem**."

- The outcome of the one year proposal should be a compelling plan for a center (follow on proposal) for "**the long-term coordinated community-driven development directions necessary to sustain the HPC tools ecosystem.**"

- "Further, many collaborative efforts lose steam when internal priority directions outweigh collaborative gain. Our approach **increases the collaborative gain by bringing together communities to address challenges caused by their dependencies.**"

OAK RIDGE National Laboratory

S4PST

Keita Teranishi, Group Leader
Programming Systems

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

OAK RIDGE National Laboratory

Team

Keita Teranishi (PI, ORNL)

Pedro Valero Lara (co-PI, ORNL)

William Godoy (co-PI, ORNL)

Hartmut Kaiser (LSU)

Patrick Diehl (LSU)

Alan Edelman (MIT)

- Sameer Shende (U Oregon)
- Sunita Chandrasekaran (U Delaware)
- Michel Schanen (ANL)
- Damian Rouson (LBL)
- Johannes Blaschke (LBL)
- Ignacio Laguna (LLNL)
- Todd Gamblin (LLNL)

- Christian Trott (SNL)
- Damien Lebrun-Grandie (ORNL)
- Pat McCormick (LANL)
- Johannes Doerfert (LANL)
- Philippe Pebay (NGA)
- Francesco Rizzi (NGA)
- Hartwig Rizzi (UTK)
- Valentine Churavy(MIT)
- Suzanne Parete-Koon (ORNL)

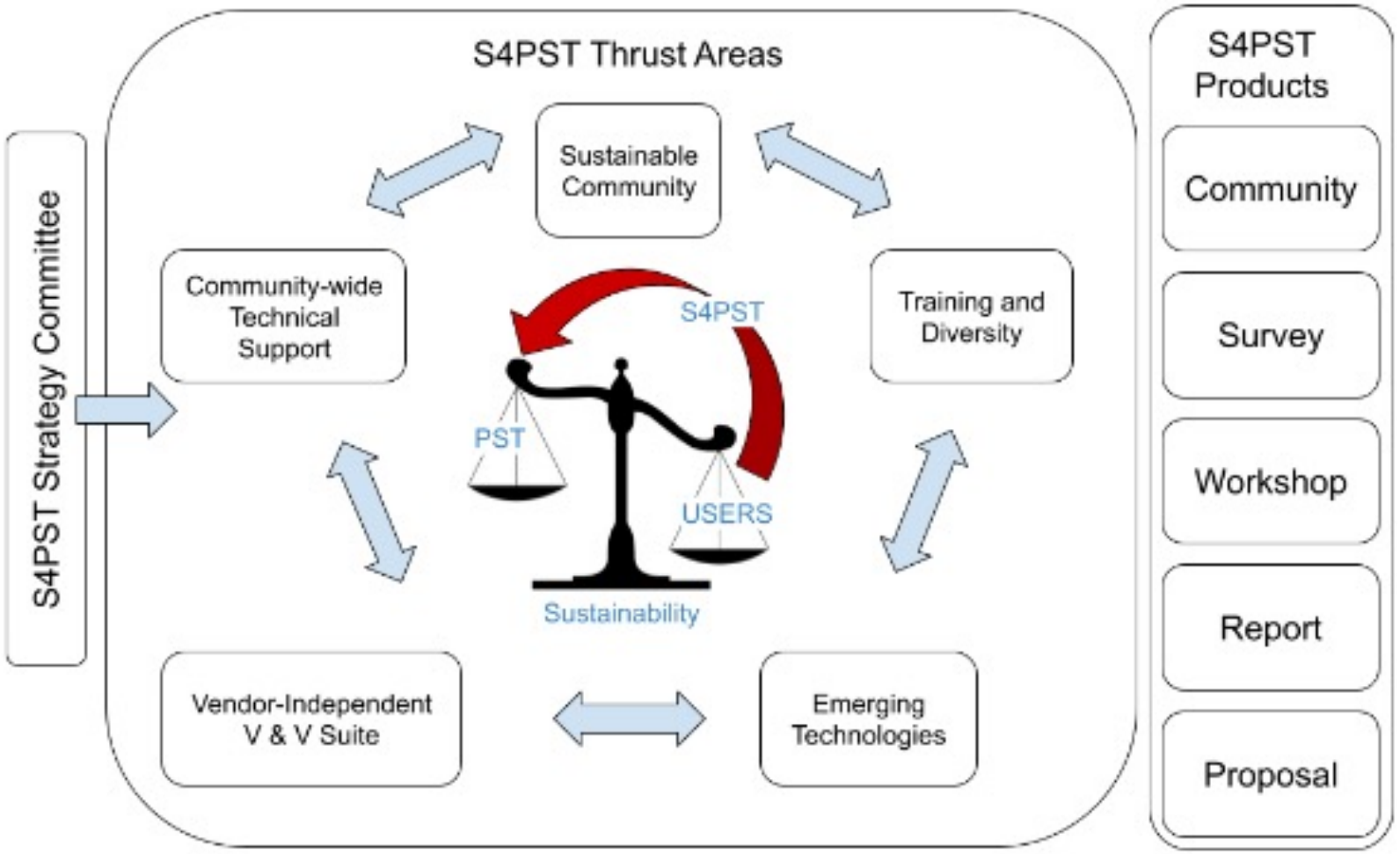
S4PST: Motivation

- ECP has created software ecosystems for scientific HPC community.

- These ecosystems can be fragile without robust programming systems and compilers.
  - Offloading ecosystem to third party or non-HPC.
  - Patchworked ecosystem with external dependencies
  - Emerging languages have seamless ecosystem model
  - Reactive approach is costly

- **The sustainability efforts should proactively address:**
  - **Social aspects (community, training, inclusiveness)**
  - **Economic aspects (time, total cost, amortization)**
  - **Technical (interoperability, capability, software/libraries)**

S4PST



S4PST

Sustainable Community
Community-Wide Technical Support
Training and diversity
Vendor-Independent Verification and Validation
Emerging Technologies
Identifying the coverage, common interests, non-DOE users

Brainstorming:

- Additional members from Sandia, LANL, ORNL, UTK, NexGen
  - Analytics

Open Scientific Software Foundation

Creating a sustainable future for scientific software

What is the foundation's mission?

Provide an equitable and resilient center of gravity that enables and supports a thriving and sustainable ecosystem of open source scientific software projects.

What is the foundation's mission?

Provide an equitable and resilient center of gravity that enables and supports a thriving and sustainable ecosystem of open source scientific software projects.

```
inputs: people, scientific_software
```

What is the foundation's mission?

Provide an equitable and resilient center of gravity that enables and supports a thriving and sustainable ecosystem of open source scientific software projects.

```
def foundation(people, scientific_software):
    ensure_future(sustainability)
```

Why a foundation?

- A foundation will help support a community that is invested, works toward common goals, shares important ideas, and provides common services and infrastructure that benefit the whole community
- A community working this way can result in something that is greater than the sum of the individual parts
- Foundations that support open source software have been a highly successful model
  - Linux, Eclipse, Apache, NumFOCUS
  - Others at https://opensource.com/resources/organizations

What the foundation is not

- Tied to or aligned with any particular sector, organization, or group
- Trying to control how projects or programs organize or operate
- Creating a one-size-fits-all approach
- Replacing other software foundations

How does this benefit ECP?

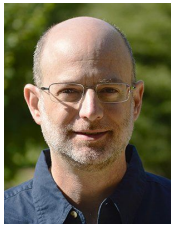
- ECP already has a strong governance model and a focus on a software ecosystem and delivery, but needs a long-term sustainability strategy
- Projects could continue to operate under a similar model within an “ECP Program” or some projects might align better with other communities, e.g. within a “Programming Systems & Tools Program”
- In either case, the foundation would make it easier and cheaper for the projects by focusing on sustainability, consolidating common services, and provide training and support for better practices
- Allows separation of sustainability concerns from new R&D efforts

Our Team

Greg Watson, Oak Ridge National Laboratory

Addi Thakur Malviya, Oak Ridge National Laboratory

Daniel S. Katz, University of Illinois Urbana Champaign



Dana Robinson, The HDF Group

Elaine Raybourn, Sandia National Laboratories

Bill Hoffman, Kitware, Inc.

John Kellerman, The Eclipse Foundation, Inc.

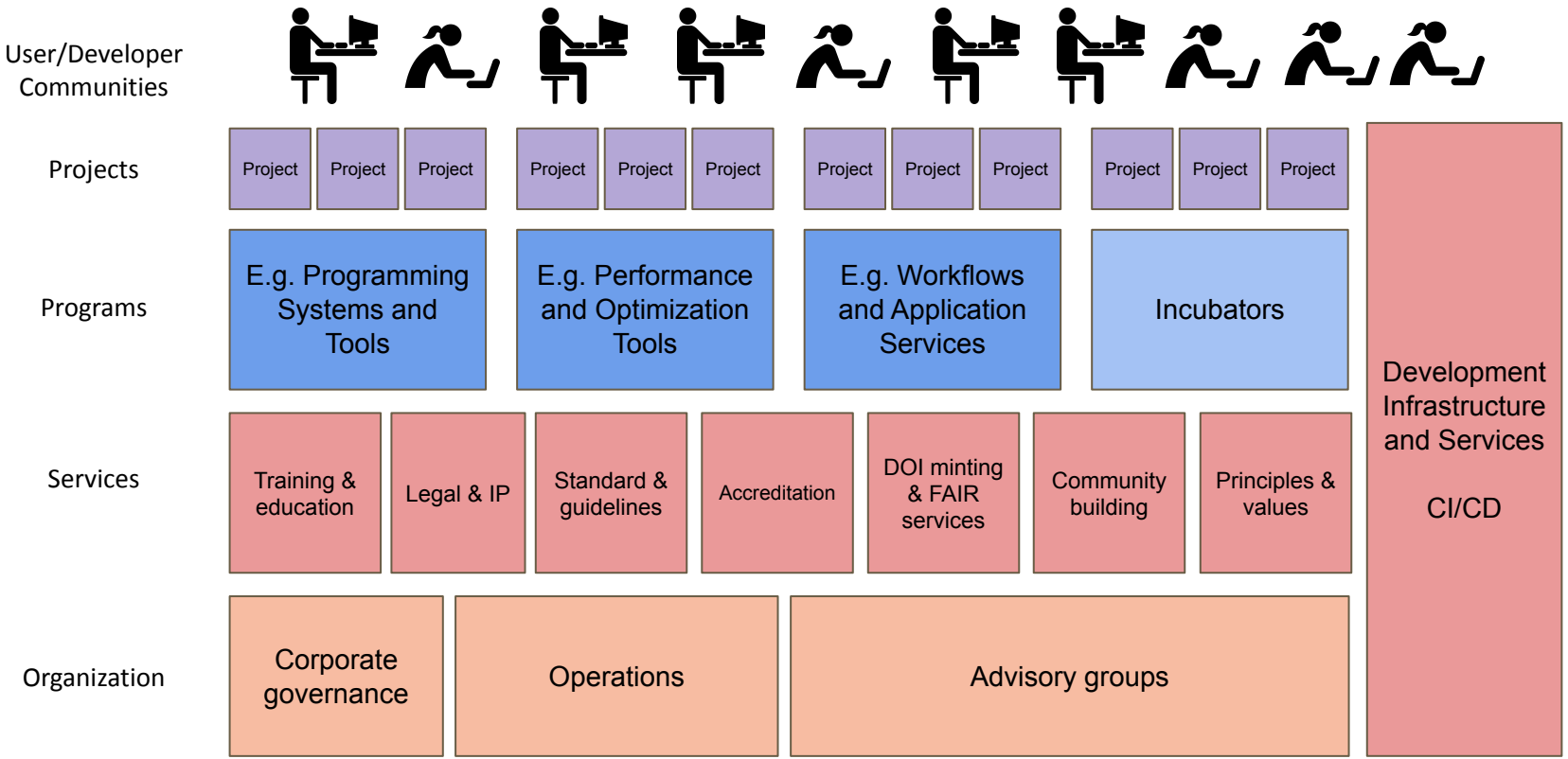
Clark Roundy, The Eclipse Foundation, Inc.

Thank you!

We welcome your feedback and involvement!

Additional Material

How might the foundation be structured?



Key concepts

User/developer communities are the keystone of the organization

Projects are centered around one or more software artifacts

Programs provide a way for sets of user and developer communities to self organize and determine the best way of operating for their community

Services are available to all projects that are part of the foundation

Advisory boards provide guidance and recommendations to the foundation

Operations provides responsibility for the day-to-day operation of the foundation, negotiations with sponsors/vendors, and facilitating meetings/events

Corporate governance is ensuring that the foundation meets the needs of the members and stakeholders and works towards long term sustainability goals

What value would we bring to software ecosystems?

- Representative organization that can advocate on behalf of the member projects
- Strong sense of affinity and attachment to other member projects, of “belonging”, of shared purpose
- Powerful position for developing vendor and stakeholder relationships (e.g. AWS, commercial development tools, etc.)
- Common infrastructure services (CI/CD, build, communication, DOIs, FAIR data, etc.)
- Well researched software engineering guidance and practices (standards, guidelines, badging, incentives, etc.)
- Training and education services relating to scientific software development
- Working with the research software engineering community to increase recognition and create educational pathways
- Intellectual property and legal services
- Reduce the costs to DOE by creating a mechanism to bring together resources from multiple funding agencies, organizations, and individuals
- Host meetings and conferences to bring people together to share ideas, particularly across scientific domains
- Provide incentives for improving software practices