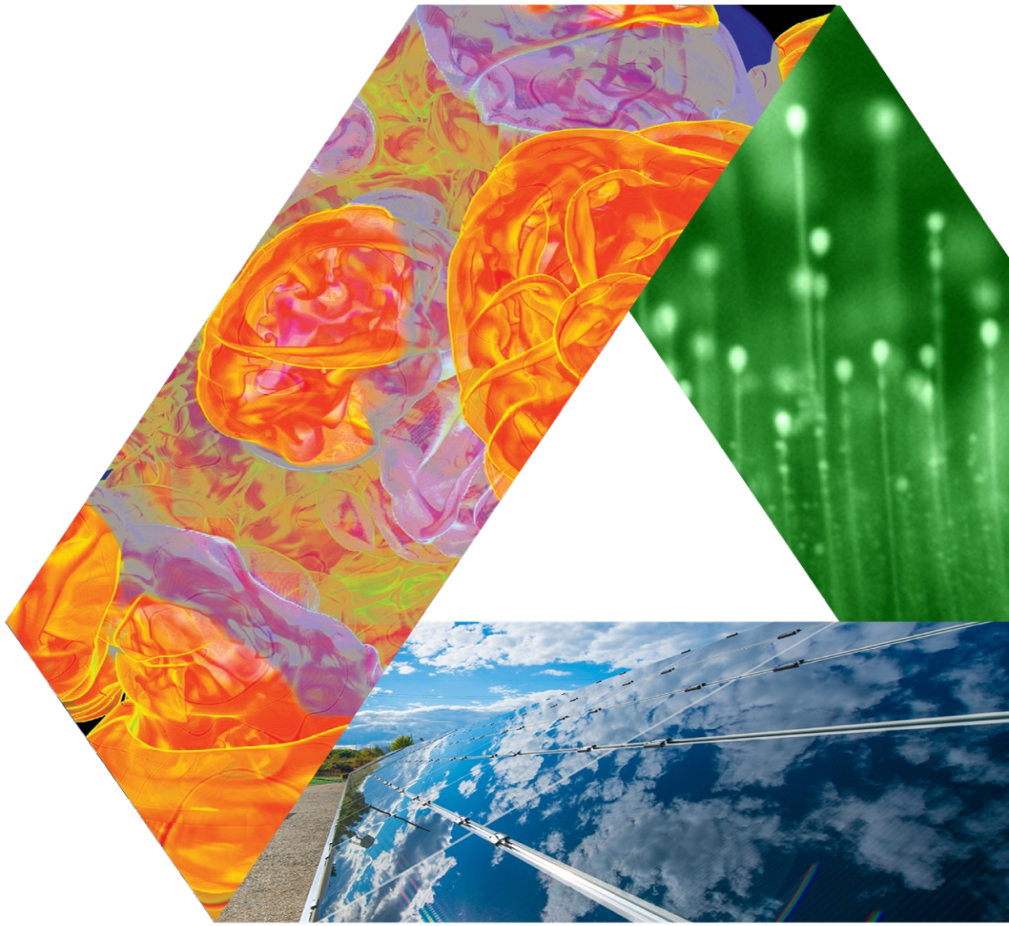
COLABS: Collaboration for Better Software (for Science)

MCS Seminar May 30, 2023

Presenter: Anshu Dubey

Contributors: David Bernholdt, Dan Gunter, Kevin Harms, Bronson Messer, John MacAuley, and Richard Gerber

https://colabs-science.github.io



Argonne National Laboratory

acknowledgements

This work was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR) at Argonne National Laboratory, Oak Ridge National Laboratory, and Lawrence Berkeley National Laboratory

Overview ASCR Sustainability Initiative

- DOE-ASCR is concerned with post ECP sustainability of software
  - Huge investments
  - Change in perception
  - Explicit instead of implicit
- Intent is to get a broad community input
- Funded six seedlings for planning sustainability organizations
  - Three seedlings have specific community focus
  - Three take a broader view

COLABS

OSSF

PESO

S4PST

STEP

SWAS

Some or all may be funded to form organizations

https://lssw.io/SeedProjects.html

Broader View Seedlings

Common Sustainability Services:

- infrastructure (such as CI/CD)
- software engineering training and education
- standards and guidelines

COLABS: Collaboration for Better Software (for Science)

- Matrixed RSEs providing services above
- Services at different granularity, include embedding personnel

PESO: Toward a Post-ECP Software-Sustainability Organization

- ❑ Libraries and tools
- ❑ Project based approach – conduit of funding to projects

OSSF: Open Scientific Software Foundation

- ❑ Additional Services
  - ❑ Marketing/sponsorship/fundraising activities
  - ❑ Sustainability funding for projects
  - ❑ FAIR data and DOI minting
  - ❑ badging and incentives
  - ❑ legal and IP advice

Community Focused Seedlings

S4PST : Sustainability for Node Level Programming Systems and Tools

- Coordinated ecosystem for compilers, abstractions, libraries, and tools for node level programming
- Similar to PESO but for a specific community

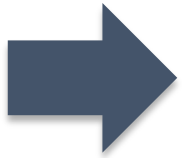
STEP : Sustainable Tools Ecosystem Project

- Co-design with hardware vendors
- Participating projects directly funded – participate in co-design with STEP

SWAS – Center for Sustaining Workflows and Application Services

- Software and services for complex workflows
- Similar to STEP -- projects directly funded

Good scientific process
requires
good software practices



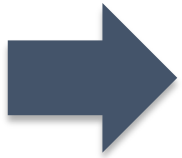
Good software practices
increase
software sustainability

Good software practices
increase
scientific productivity

Software sustainability
increases
scientific productivity

Argonne NATIONAL LABORATORY

Good scientific process
requires
good software practices



Good software practices
increase
software sustainability

Good software practices
increase
scientific productivity

Software sustainability
increases
scientific productivity

Science with Computing is at best as
good as the software that produces it

Argonne NATIONAL LABORATORY

COLABS

Comprehensive services in support of scientific software stewardship and sustainability

COLABS

Comprehensive services in support of scientific software stewardship and sustainability

Elevate scientific software development through

- Availability of well-trained professionals
- Extensive training
- Education
- Community-building efforts
- Advocacy for both better software and for the people responsible for it

A modest research component on the “science of scientific software”

- In the spirit of SSSDU workshop report
- To help improve COLABS services and training and the broader community.

The Research Software Engineer role

People with many job titles contribute to the development and maintenance of scientific software

Software professional

Computer Scientist

Domain scientist

Computational Scientist

Any of them can assume full or part time role of research software engineer

The Research Software Engineer role

People with many job titles contribute to the development and maintenance of scientific software

Software professional

Computer Scientist

Domain scientist

Computational Scientist

COLABS adoption of term RSE

- ..As part of a world-wide movement which has settled on the term "research software engineer" (RSE) for this role
  - "those who regularly use expertise in programming to advance research" -- US-RSE (professional society)
- Many research-oriented institutions have a hard time valuing contributions to software development and sustainment
  - This applies to full and part time RSE's
- COLABS will advocate for better recognition and rewards for software-related contributions with member institutions and sponsors

Any of them can assume full or part time role of research software engineer

The Research Software Engineer role

People with many job titles contribute to the development and maintenance of scientific software

Software professional

Computer Scientist

Domain scientist

Computational Scientist

COLABS adoption of term RSE

- ..As part of a world-wide movement which has settled on the term "research software engineer" (RSE) for this role
  - "those who regularly use expertise in programming to advance research" -- US-RSE (professional society)
- Many research-oriented institutions have a hard time valuing contributions to software development and sustainment
  - This applies to full and part time RSE's
- COLABS will advocate for better recognition and rewards for software-related contributions with member institutions and sponsors

- Project staff can become members of COLABS for negotiated fractions of their time
- COLABS will also have a pool of full time staff for sustainability services in its catalog
- Project staff in COLABS will have the option of joining the pool of service providers

Any of them can assume full or part time role of research software engineer

COLABS Activities

A Partial Catalog of Sustainability Services

Central Services

- Services that are more effectively centralized
- Setup and maintenance of CI/CD pipelines
- Building and maintaining containers

Essential Services

- Baseline levels of effort
- Specific activities determined with help from COLABS

Advanced Services

- Above baseline levels
- Porting, test development, refactoring, etc.
- May include lightweight proposal process

Coaching

- Help with assessment and improvement of software practices.
- Use COLABS training to help clients in enhancing sustainability

Subject Matter Experts

Pool of experts needed occasionally by projects

- Test design
- UI/UX design
- Security protocols
- Licensing etc.

COLABS Activities - Central Services

- Some software engineering activities are hard for individual projects to implement effectively because...
  - The details are finicky and change frequently
  - Implementation in national lab/facility environments may impose additional requirements
  - Etc.
- CI/CD pipelines are an obvious example
  - Tools and best practices for implementation change frequently
  - Specialty environments (e.g., computing facilities) and (more) secure environments (e.g., local hardware in national labs)
- Containerization is another example
  - Various tools, formats, limitations/restrictions for various flavors of containers
  - Different facilities, cloud services, etc. support different tools
  - Ensuring ability to effectively utilize specialize hardware (e.g., accelerators, interconnects)
- COLABS expects to have staff with high levels of knowledge and experience in such areas and can more efficiently support such needs for projects
  - Option to train project staff more deeply

COLABS Activities

Training
- To COLABS members and client projects
- Larger community

Community of Practice
- Share experience and expertise
- Professional growth
- Engagement with larger community and advocacy for software professionals

COLABS Activities

Workforce Development : goal of building scientific software development workforce

- Developing and delivering educational programs
  - Internships and co-op opportunities
- Bringing practical scientific software engineering knowledge and experience into university programs
- Special attention to underrepresented groups
- Outreach to colleges and universities that have had little or no connections to Scientific Computing

COLABS Activities – Special Cases

Incubation

- Help with initial setup and development
- Training project personnel
- Help bring the project to a self-sufficient operational level

Assuming Maintenance

- Mature project without enough resources
- Of value to community
- With guidance sponsors or community
- Funding options
  - Some funding set aside by COLABS
  - Sponsors provide necessary funding
  - Users contribute efforts to COLABS for project maintenance

Working with COLABS

- Embedded project staff
  - COLABS can provide financial support and training
  - Advocacy
- Projects in need of help (or unaware that they need help)
  - COLABS can provide needed services
- Pool of expertise that most projects need only occasionally
  - Get the help without having to hire someone

Working with COLABS

- Embedded project staff
  - COLABS can provide support and training
  - Advocacy
- Projects in need of help (or unaware that they need help)
  - COLABS can provide needed services
- Pool of expertise that most projects need only occasionally
  - Get the help without having to hire someone

Benefits to COLABS Members

- Professional development and growth opportunities
- Training to stay current with the state-of-the-art
- Access to a community of like-minded people
- Appropriate recognition of contributions to COLABS clients
- Advocacy for recognition and career path
- Greater (funding) stability for staff in RSE roles

Working with COLABS

- Embedded project staff
  - COLABS can provide support and training
  - Advocacy
- Projects in need of help (or unaware that they need help)
  - COLABS can provide needed services
- Pool of expertise that most projects need only occasionally
  - Get the help without having to hire someone

Benefits to COLABS Members

- Professional development and growth opportunities
- Training to stay current with the state-of-the-art
- Access to a community of like-minded people
- Appropriate recognition of contributions to COLABS clients
- Advocacy for recognition and career path
- Greater (funding) stability for staff in RSE roles

Benefits to Project (PI)

- Greater stability for services critical for project
  - Less need to cut corners on quality and user-engagement
  - Some continuity in face of erratic funding cycles
- Support for capabilities that can be hard to sustain (especially for smaller teams)
- More diversified expertise available when needed

Benefits of COLABS for Sponsors

- Increased attention to and action on software sustainability
- Economies of scale in sustainability services
  - The design of COLABS can support the needs of multiple sponsors
  - ASCR is the initial sponsor
  - Can imagine other DOE offices funding COLABS to support application projects, SciDAC projects, experimental and observational user facilities, etc.
- Better match of tasks and expertise for more efficiency
- Highly trained workforce
- More professional development, career support --> greater retention
  - COLABS also plans a significant workforce development component to build a stronger, deeper pipeline for growing needs for people in the RSE role
- Possibility of bridging funding gaps for useful software

Still Under Discussion

Allocation of COLABS Resources

- Training available to all, to promote better software practices
- Sponsors engagement in defining a "constituency" of projects that we're "allowed" to support at all levels
- Basic services available to all projects in the constituency
- Coaching available to all projects in the constituency, to help implement better software practices
- Allocation for higher service levels is under discussion

Still Under Discussion

Allocation of COLABS Resources

- Training available to all, to promote better software practices
- Sponsors engagement in defining a "constituency" of projects that we're "allowed" to support at all levels
- Basic services available to all projects in the constituency
- Coaching available to all projects in the constituency, to help implement better software practices
- Allocation for higher service levels is under discussion

Engagement with Industry, Other Sponsor and Projects

- Companies and other agencies could buy into COLABS as sponsors
- COLABS can support industry/other agency projects with sponsor concurrence
  - e.g., DOE currently engages with other agencies as a leading federal agency in HPC/scientific computing; could do the same for scientific software stewardship
- COLABS can collaborate with other software sustainability organizations
- Potentially federation of SSOs

Workforce Development and Research

- We believe more people are needed in RSE role – plan for significant workforce development activity to help address that need
  - Engage with teaching institutions to incorporate practical scientific software development and engineering into their curricula
  - Complement with internship and co-op programs to give students hands-on experience
- Emphasize inclusion of people from underrepresented and marginalized groups in education, internship, co-op efforts
  - Reach out to minority serving institutions and those that have no or limited connection to HPC
- COLABS staff will be encouraged to serve as instructors, mentors

Workforce Development and Research

- We believe more people are needed in RSE role – plan for significant workforce development activity to help address that need
  - Engage with teaching institutions to incorporate practical scientific software development and engineering into their curricula
  - Complement with internship and co-op programs to give students hands-on experience
- Emphasize inclusion of people from underrepresented and marginalized groups in education, internship, co-op efforts
  - Reach out to minority serving institutions and those that have no or limited connection to HPC
- COLABS staff will be encouraged to serve as instructors, mentors

- A modest R&D activity with the goal of improving COLABS services and training
  - In the spirit of the findings of the Workshop on the Science of Scientific Software Development and Use (SSSDU) report to be published soon
- Involvement in software science R&D is another professional development opportunity for COLABS team members

Brochure for SSSDU available

COLABS Design is Scalable in Multiple Dimensions

Organizational scalability

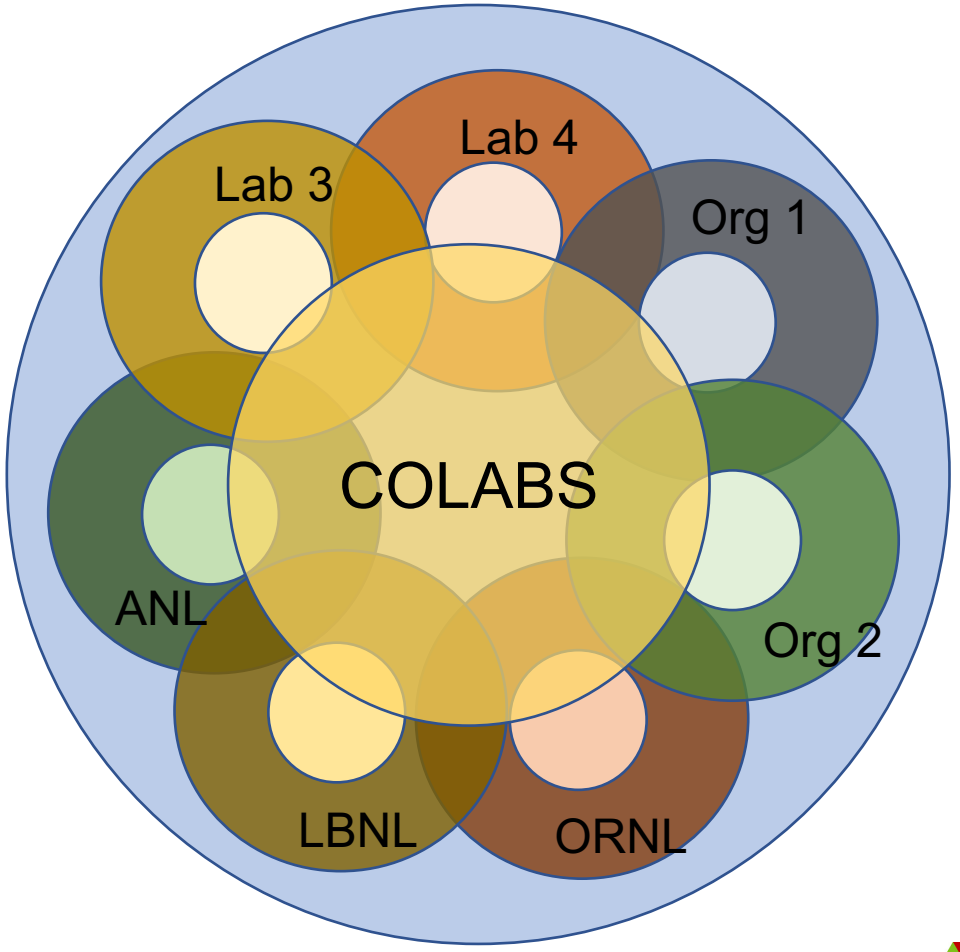
- We envision an institutional footprint for COLABS which is similar to that of the constituency it is meant to serve
- Tempered by the desire to maintain critical mass for communities of practice to be meaningful at the institutional level

Budgetary scalability

- The services COLABS provides can be scaled to the budget available (within limits, of course) -- more funding --> more services

Scalability to additional sponsors

- The same structure and governance model, and to a large extent the same (types of) people can support multiple sponsors
- ASCR is first sponsor. Others could be Office of Science offices, user facilities, DOE applied offices, etc.



COLABS Funding Model – Not Completely Known

- Stable funding obtained directly, indirectly, or a mix of the two
- Directly
  - From the sponsor to COLABS as block funding
  - Allocation of resources to projects based on evaluation of project needs and possible sponsor guidance
- Indirectly
  - PI can specify services they want from COLABS, and sponsor can provide funding to COLABS earmarked for the PI's project
- Project can buy additional services from COLABS directly
  - e.g., if they like the COLABS approach and sponsor support doesn't fully address needs
- Other ….