COLABS: Collaboration of ORNL, LBNL, and ANL for Better Software

Investigator	*Institution*	*Role*
Anshu Dubey	**Argonne National Laboratory**	**Lead PI**
Kevin Harms	Argonne National Laboratory	
Dan Gunter	**Lawrence Berkeley National Laboratory**	**Co-PI**
Richard Gerber	Lawrence Berkeley National Laboratory	
John Macauley	Lawrence Berkeley National Laboratory	
David E. Bernholdt	**Oak Ridge National Laboratory**	**Co-PI**
Bronson Messer	Oak Ridge National Laboratory	

ABSTRACT

The Exascale Computing Project (ECP) has brought the issue of scientific software sustainability into sharp focus with a clear indication that carving out support for sustainability from research-based software funding is no longer tenable. The bulk of ECP funding supported activities that have in recent years become recognized as work done by research software engineers (RSEs). These experiences guide our conceptualization of a distributed software sustainability organization (SSO) to support scientific software sustainability and stewardship needs for the Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR), but also extendible to other sponsors. The SSO will be staffed primarily by RSEs who are highly trained and experienced in the craft of software development. The proposed organization, the Collaboration of ORNL, LBNL, and ANL for Better Software (COLABS) SSO, will initially be anchored at Argonne National Laboratory (ANL), Lawrence Berkeley National Laboratory (LBNL), and Oak Ridge National Laboratory (ORNL). Our governance, engagement, and budgetary models are conceptualized to be scalable along multiple dimensions to include additional members joining the SSO, and engagement with other sponsors and organizations with similar objectives. We believe that this flexibility will allow the SSO to successfully address the software sustainability and stewardship needs of a substantial portion of the DOE scientific software base. An additional benefit of our approach will be a well-trained and diverse workforce of scientific software developers.

The COLABS software sustainability organization (COLABS SSO) will provide a wide range of services to client software projects and the broader community in partnership with ASCR's user facilities: training and outreach, essential and advanced software engineering services, along with a modest research and development (R&D) effort. The R&D component will ensure continued enrichment of the SSO's offerings. The SSO will place a strong emphasis on workforce development and retention and will provide the long-term stability, training, and support to enable and encourage RSEs and other staff to build their careers and excel in this role. We anticipate that many software developers currently contributing to the ECP would be excellent candidates for the initial cadre of SSO staff.

During the seedling phase of the SSO development, covered by this proposal, we will reach out to the community to introduce our approach and gather input to further refine our plans, work with ASCR to better understand their expectations of how the SSO will operate and to flesh out key concepts in the design, including the governance model, management approach, and budgetary model. Finally, once appropriate details are known, we can finalize the team for the operational SSO and begin recruiting staff.

If multiple SSOs are funded by ASCR, we intend to work cooperatively to maximize the value provided to the sponsor and client projects while minimizing redundancies.

1. Narrative

1.1 Objectives

The Exascale Computing Project (ECP) has brought the issue of scientific software sustainability into sharp focus. It has become apparent that the dependence of the Department of Energy (DOE) scientific enterprise has reached a state where carving out support for sustainability from research-based software funding is no longer tenable. The bulk of ECP funding supported activities that have in recent years become recognized as work done by research software engineers (RSEs)[*]. It is equally apparent that such support must continue to protect not only the ECP investment but, eventually, the entire DOE scientific investment. A universal challenge faced by the teams engaged in ECP work was the recruitment and retention of an appropriately trained workforce. These experiences guide our proposal to conceptualize and build a distributed software sustainability organization (SSO) to support scientific software sustainability and stewardship needs for the DOE Office of Advanced Scientific Computing Research (ASCR), but also extendible to other sponsors. The SSO will be staffed primarily by RSEs who are highly trained and experienced in the craft of software development. The proposed organization, the Collaboration of ORNL, LBNL, and ANL for Better Software (COLABS) SSO, will initially be anchored at Argonne National Laboratory (ANL), Lawrence Berkeley National Laboratory (LBNL), and Oak Ridge National Laboratory (ORNL). Our governance, engagement, and budgetary models are conceptualized to be scalable along multiple dimensions to include additional members joining the SSO, and engagement with other sponsors and organizations with similar objectives.

This proposal is for what ASCR has referred to as a "seedling" phase with an extremely limited budget and period of performance ($125,000 for one year), meant to provide an opportunity to work with ASCR and the other funded seedling projects to refine the general vision for DOE software ecosystem stewardship and, our approach in particular. The bulk of the proposal (Sec. 1.3–1.5) expands on our current conception of an SSO designed to meet the objectives above. Our objectives during the brief seedling phase are outlined below and discussed in greater detail in Sec. 1.6.

- Broad discussions to introduce the COLABS SSO approach to the constituency[†] and to other potential sponsors and gathering of input to help refine the design.
- Discussions with ASCR to better define the "bounding box" for the operational SSOs.
- Coordination with other seedlings to minimize redundancies and identify areas of collaboration.
- Fleshing out the governance model and management approach, and developing a budgetary model for the COLABS software sustainability organization (COLABS SSO).
- Fleshing out the team of institutions and individuals that will comprise the initial operational SSO, and when appropriate, beginning to recruit individual staff members.

1.2 Current practice and its limits

Scientific software is, increasingly, the backbone of the entire science mission of the DOE. Computation and data have become a critical part of scientific workflows, and scientific advancement relies on both the capabilities and the correctness of the underlying software. Software often has long-term value beyond the publications it produced and the project that created it. As a result, software stewardship and sustainability

[*]The UK Society of Research Software Engineering states "An RSE combines professional software engineering expertise with an intimate understanding of research." [1]

[†]We use the term *constituency* to denote the full set of projects that the sponsor has designated for the SSO to support, and the term *client project* to denote those projects which have chosen to engage with the SSO. Generally, we expect engagement with the SSO to be optional (working with willing projects will be far more productive), and engagements with any particular project may not necessarily be continuous (according to their needs and interests). So at any given moment, the client projects are likely to be a subset of the constituency.

have become critical needs for the entirety of the scientific enterprise. Outside of the ECP most important decisions about software stewardship continue to be made in the margins of research funding cycles, by individual project teams, and with little oversight or recognition from their line or project management. The ECP has put a spotlight on the topic of software sustainability beyond the research phase, and also on protecting investments in software development.

As the ECP has demonstrated, creating reliable, high-quality software requires the attention and discipline of modern software engineering best practices. In particular, the entire software development lifecycle (SDLC) – requirements gathering, design, development, testing, deployment, and user support and maintenance – needs to be addressed. The common practice of relying upon post-doctoral fellows and scientific staff to address these issues without appropriate knowledge or training in software engineering, and with different incentives, is likely to produce substandard software in addition to being inefficient.

Support for career development and training for software engineering is inconsistent across the national laboratories and other participants in the DOE scientific software ecosystem. This makes the career path for RSEs confusing and difficult to navigate, discouraging recruitment and retention. While there is general support for attending conferences and enrolling in professional development opportunities, there is little sense of a baseline set of skills and knowledge needed to be effective as an RSE in scientific computing.

Taken together, these factors suggest that the status quo will result in scientific software being increasingly more difficult and costly to sustain and use, and in experts who can improve the situation remaining in short supply. The COLABS SSO approach will address these challenges by bringing knowledge of skilled practitioners of software engineering to support projects in the scientific software ecosystem, providing services where applicable, and training in such practices to an even larger audience.

1.3 Proposed COLABS SSO approach

This section, as well as Sec. 1.4–1.5, describes our current conception of the COLABS SSO. Sec. 1.6 describes the activities we plan to carry out during the "seedling" phase, covered by this proposal, to further refine our design and begin realizing it.

The COLABS SSO will provide a wide range of services to client software projects and the broader community in partnership with ASCR's user facilities: training and outreach (Sec. 1.3.1), essential and advanced software engineering services (Sec. 1.3.2–1.3.3), along with a modest research and development (R&D) effort (Sec. 1.3.4). The R&D component will ensure continued enrichment of the SSO's offerings (e.g., use of artificial intelligence (AI) techniques to analyze repository data or studying testing archives to refine corresponding best practices). A carefully designed management structure (Sec. 1.3.5), along with a flexible governance model (Sec. 1.3.6) will oversee the services and guide the allocation of resources to client projects. The SSO will place a strong emphasis on workforce development and retention and will provide the long-term stability, training, and support to enable and encourage RSEs and other staff to build their careers and excel in this role (Sec. 1.3.7). We anticipate that many software developers currently contributing to the ECP would be excellent candidates for the initial cadre of SSO staff.

We envision the SSO as a multi-institutional distributed organization that will initially be anchored at ANL, LBNL, and ORNL, but, given adequate funding levels, can expand to include other national laboratories, academic institutions, and possibly commercial entities as shown in Figure 1.1.

The COLABS SSO concept is flexible in that the same structure and approach can be used to serve not only ASCR, but additional sponsors as well. Other offices within DOE can potentially engage the COLABS SSO to support their own portfolios, for example, ECP Application Development (AD) or Scientific Discovery through Advanced Computing (SciDAC) applications, software stacks for experimental and observational user facilities, etc. Likewise, SSO member institutions may wish to leverage the SSO's approach, experience, and capabilities to support software of strategic interest as well (e.g., developed under internal programs or for non-DOE sponsors).

The COLABS SSO concept is likewise flexible concerning how client software projects engage with

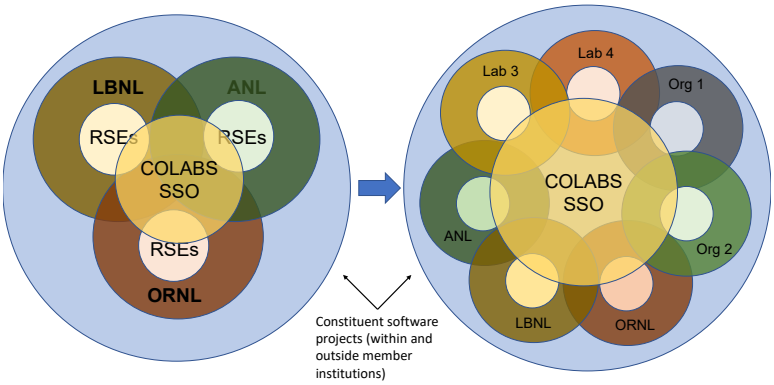

Figure 1.1: *The initial organization of the COLABS SSO (left) and a growth path to including more institutions as members (right). The SSO draws on RSEs and other staff at participating institutions to serve projects in the constituency (outer blue circle), which may be based at many institutions. Additional member institutions can be added following the same model, and the new footprint of the SSO will cover more of the constituency.*

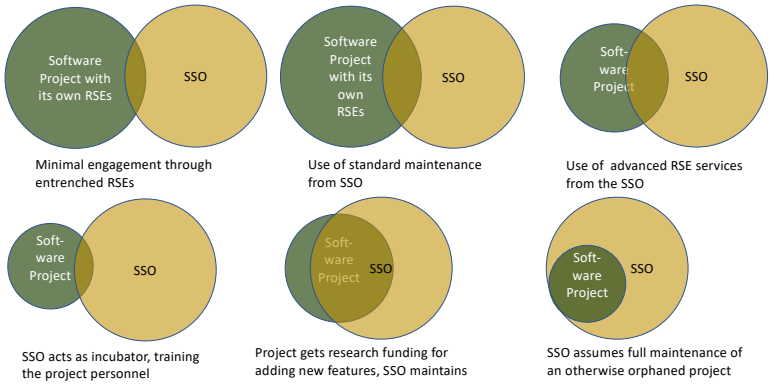

Figure 1.2: *Engagement scenarios for software projects interacting with the SSO*

it. We expect that many projects will engage over an extended period, to the extent that the SSO RSEs supporting them are effectively embedded as members of the project team. In addition to providing services directly, the SSO RSEs will also coach the team members in software process improvement to increase the team's own capabilities. Some projects already have their own RSEs who could spend part of their time on the SSO's payroll performing sustainability-related activities, but also taking training and sharing knowledge and experience with others, as expected of SSO staff. And in some cases, the SSO may take over complete responsibility for the sustainment of critical software which is no longer being maintained.[‡] Fig. 1.2 illustrates several engagement scenarios.

1.3.1 Training and outreach

An essential part of the COLABS SSO concept is advocacy for software as a first-class product of research and providing all scientific software developers with the training and information needed to improve their software development processes and produce more sustainable software. The Training and Outreach component of the SSO will also be responsible for other aspects of workforce development and retention, including pipeline-building efforts with universities and community colleges, and RSE-focused internship programs.

[‡]Though a more likely scenario is that the SSO would help the project transition to an alternative, to replace the orphaned dependency.

Outreach: elevating scientific software and its developers. We believe that the most effective way to achieve these objectives is to help continue and accelerate the cultural change we have begun to see with formal recognition of the RSE role. Examples of the kinds of activities we envision in this area come from the Interoperable Design of Extreme-Scale Application Software (IDEAS) project, in which principal investigators (PIs) Bernholdt and Dubey participate. They include the organization of events at conferences that facilitate discussions about software in science [2]. IDEAS project members have also (co-)organized workshops for DOE program managers [3] and for the inter-agency Networking and Information Technology Research and Development (NITRD) program [4]. Our Better Scientific Software Fellowship program [5] has been extremely valuable for raising awareness and bringing new ideas and perspectives on software development to the DOE and National Science Foundation (NSF) communities.

Training software developers. There is a general lack of needed software engineering training among the developers of scientific software. Efforts such as INovative Training Enabled by a Research Software Engineering Community of Trainers (INTERSECT) [6], as well as tutorials [7], webinars [8] and web resources [9] hosted by the IDEAS project are beginning to address this gap. The leadership of the proposed SSO is engaged with, and in many instances, leading these efforts. The COLABS SSO will leverage the connections and experience from these efforts, as well as organizational training expertise, such as the Argonne Leadership Institute [10] to build a comprehensive training program. With an effective training program in place, we can recruit talent based on potential rather than prior knowledge and experience.

While primarily targeting client software projects and SSO staff, training and informational resources will be made available more broadly to promote awareness and improvement in software development practices in the larger community. Advanced training would provide professional development opportunities for RSEs as they proceed along a career track and would easily be made available more broadly as well.

Filling the pipeline: student internships and educational partnerships. We recognize that workforce development and retention is a general problem for DOE at present, however, the RSE workforce poses particular challenges. The skills we require are also valued by the "tech" industry and they are able and willing to pay higher salaries than the national labs or academia. We propose to address this using a two-pronged approach. One is to promote the inclusion of relevant coursework in college and university curricula to broaden the talent pool. Where it is useful, COLABS staff may offer to teach relevant classes or modules. The second approach is to institute strong internship and practicum programs for students at the graduate, undergraduate, and community college levels to improve their skills in *scientific* software engineering, and to expose them to practical experience. The experiential component will have the benefit of giving students a taste of how they can actively contribute to the scientific enterprise, often cited as a non-monetary benefit in recruiting for national laboratories and academia.

Our educational engagements will also be an important aspect of the SSO's inclusivity, diversity, equity, and accessibility (IDEA) strategy. We will emphasize minority serving institutions (MSIs) and Historically Black Colleges and Universities (HBCUs) as partners, and individuals from underrepresented groups in the recruitment and selection process for internship and practicum programs.

1.3.2 Essential services

Complementing the more general training activities, the SSO will provide a broad array of software engineering-related services to client projects. Some services will be more scalable and efficient if they are supported centrally, while others will be provided directly to individual projects. From a management perspective, Essential Services can be viewed as a level of RSE effort allocated to serve the basic needs of the constituency. How this effort is used by client projects is guided by the needs of that project, within overall bounds set by the SSO's governance model (see Sec. 1.3.6).

Centralized support services. Services in this category are complex and time-consuming to set up and manage and are needed by many projects, and therefore are more efficiently handled by SSO staff. Perhaps

the most common example is continuous integration (CI) pipeline setup and maintenance. This is particularly challenging when using computing resources where the environments are specialized, (nearly) unique, and likely to be complicated by security and other constraints. Having a team within the SSO with the experience to set up and maintain the CI pipelines in these environments for client projects will be more scalable and efficient than each project doing it on its own. Other such services might include metrics dashboards, static and dynamic code analysis tools, and other similar tools where experienced SSO staff can help with both the configuration of the tool and the interpretation of its results.

RSE engagement with client software projects. In addition to the centralized services mentioned above, we will engage with projects on an individual basis to help identify and implement the changes that will most benefit them at any given time. This would involve an assessment of the current status of the project's software development practices, gathering input on pain points and particular needs or goals, which would inform the planning for improvements. We will advocate for continual, incremental software process improvement, following the Productivity and Sustainability Improvement Planning (PSIP) approach [11].

Essential Services activities will span a very wide range, but a few examples might include instituting peer code review practices, adopting coding standards, setting up testing for performance regressions, reproducing and triaging user support issues, or retiring technical debt.

Essential Services will both benefit from and provide a laboratory for the SSO's R&D activities (Sec. 1.3.4), through, for example, a better understanding of the benefits of various software process improvements, the application of novel tools, or implementation of community-building strategies.

1.3.3 Advanced services

A portion of the SSO's resources will be reserved for more complex and/or resource-intensive services, which we refer to as Advanced Services. Advanced Services requests will come in the form of lightweight proposals from client projects, which will be evaluated by a committee formed by the SSO to guide the allocation of these resources (see also 1.3.6).

RSE engagement with client software projects. The primary distinguishing characteristic of Advanced Services activities is that they require resources beyond what is allocated at the Essential Services level, but are more episodic than Essential Services. Examples of activities under Advanced Services may include porting to an emerging hardware architecture, transitioning to a more portable programming model, a major revamp of a test suite, or improving interoperability across several projects.

Access to high-end and emerging hardware platforms. We expect that many client projects will want access to the latest in high-end and emerging hardware platforms for purposes of porting and testing software packages. The inclusion of the ASCR user facilities (Argonne Leadership Computing Facility (ALCF) [12], Energy Sciences Network (ESnet) [13], National Energy Research Scientific Computing Center (NERSC) [14], and Oak Ridge Leadership Computing Facility (OLCF) [15]) in the proposal ensures that SSO's services dependent on the facilities (e.g., hardware resources for porting and testing, expertise on porting to emerging architectures) will be covered. The SSO will work with hosting organizations to arrange access and needed support. Where possible, we will engage SSO staff in these activities to help diffuse the knowledge.

Advanced Services will also engage with the SSO's R&D effort (Sec. 1.3.4). For example, these "one-off" kinds of services will be ripe for exploration of how they might be generalized and automated.

1.3.4 Research and development: the science of scientific software

While the focus of the SSO is on serving software development teams, we believe that some R&D activities will be important to facilitate the SSO's work. In particular, we envision a modest research effort in the spirit of the ASCR basic research needs workshop on the Science of Scientific Software Development and Use [16] to better inform the SSO's services and training, as well as benefitting the broader scientific software community. We expect that the SSO's interactions with a sizable number of scientific software projects

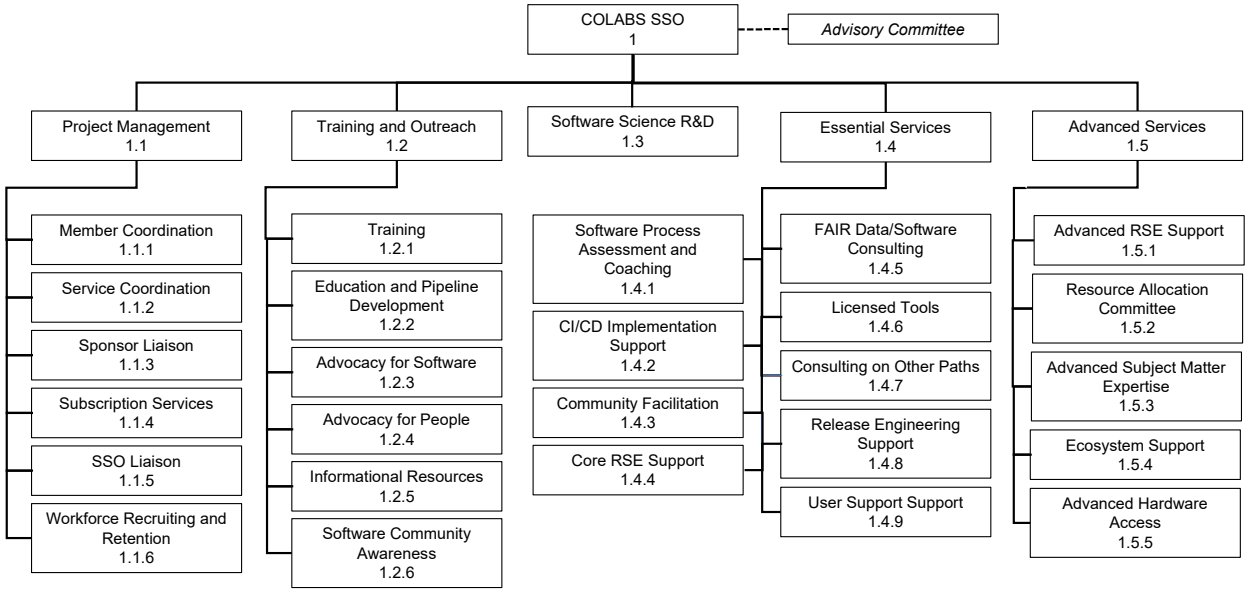

Figure 1.3: *Preliminary work breakdown structure for COLABS SSO.*

will provide a rich dataset that can use AI-based and traditional analyses to help improve offered services. The staff of the SSO need to remain cognizant of wider software engineering research as well as advances in, for example, the use of AI for code generation, and other relevant topics, to remain effective, which will be facilitated by active engagement in the software engineering research community.

1.3.5 Management

Work breakdown structure. The tentative work breakdown structure (WBS), shown in Fig. 1.3, depicts the key areas of the SSO as we currently envision it. Level 2 of the WBS corresponds primarily to the major areas of activity of the SSO plus management of the project itself. Level 3 should be considered notional at this point, but illustrates some of the major types of activity we anticipate in each area of the project. We expect the WBS, especially level 3, will be refined and revised as we progress through the seedling phase.

Leadership and decision making. The Senior Leadership Team (SLT) will consist of Dubey and Bernholdt as Co-Directors of the SSO, Gunter as Director of Research Software Engineer Staff, and a IDEA director (to be named). The Extended Leadership Team (ELT) will also include the leaders of the level 2 areas of the WBS, the institutional PIs of the SSO members, and representatives of the ASCR user facilities (ALCF [12], ESnet [13], NERSC [14], and OLCF [15]). We have used this approach in large, complex projects in the past and find that it is useful to be able to engage both technical and institutional perspectives in important discussions. The institutional representatives also provide an interface to line management for those staffing the project. The facility representatives provide an interface to the services hosted at the facilities (e.g., CI pipelines, or support for porting to new architectures). Additionally, because in some respects, the COLABS SSO will operate more like a user facility than a typical ASCR R&D project, their experience and expertise from that perspective will also be beneficial. Significant decisions will be made by consensus of the SLT based on input from the ELT.

Advisory committee. To help guide the management team, we will form an external advisory board whose members will be internationally-known advocates of software sustainability from the DOE labs, academia, industry, the international community (e.g., the Software Sustainability Institute [17]), the Software Engineering Institute (SEI) [18], and the U.S. Research Software Engineer Association (US-RSE) [19]. The advisory board will meet regularly with the ELT for updates on the SSO and provide advice. We expect 2-4

meetings per year will provide the appropriate degree of interaction.

1.3.6 Governance model

The governance model for the SSO is designed to be flexible and scalable in multiple dimensions. We provide some initial thoughts about a governance model below, with the expectation that it will be refined in conversation with the sponsors and community feedback during the seedling phase.

We believe that the sponsor should define the constituency served by the SSO. If additional sponsors wish to "buy-in" to the SSO, they define the additional constituencies they want us to support. Our conception of the SSO focuses on *sustainability* as distinct from the research and development activities that DOE has supported historically. We expect that most of the projects we work with will have R&D funding and that SSO staff will work in close collaboration with the active project team, though this is not strictly necessary from the SSO's standpoint. With sponsor approval, the SSO can assume sole responsibility for the sustainment of critical software. The line between R&D and sustainment activities can sometimes be unclear. We expect to maintain an ongoing dialog with the sponsors, even during the operation of the SSO to ensure we stay on the appropriate side of the line.

Training and Essential Services will be available to any member of the constituency upon request. Essential Services will be guided by a defined level of effort (based on the portion of the budget allocated to Essential Services and the size of the constituency), though these will be treated as soft caps, and may be adjusted dynamically by SSO management. For example, if not all of the projects in the constituency choose to engage, more resources will be available for those that do. As stated in Sec. 1.3.2, we generally expect the client projects to take ownership of the improvements made with SSO help.

As mentioned in Sec. 1.3.3, we will manage Advanced Services through a lightweight proposal process. Criteria for the evaluation will include the expected impact of the services on the project, the impact of the project itself (i.e., scientific or other impacts), the availability of people and other required resources, and a commitment by the client project to sustain the benefits obtained from the engagement.

1.3.7 Staffing and partnerships

The primary role needed to staff the SSO is that of the research software engineer who will be engaging with client projects, conducting outreach and training, and participating in the SSO's internal R&D activities. Generally, we will try to assign staff members to client projects for extended periods to allow them to become familiar with the project and team members, effectively "embedding" them with the client. Indeed, some projects already have RSEs embedded in their project. For these projects, we would propose that some of the embedded RSE's time may be used for active engagement with the SSO for mutual benefit.

The team currently developing the COLABS SSO concept is small but we envision a larger footprint in operation, as illustrated in Fig. 1.1. In practice, the institutional footprint of the SSO will need to balance between trying to cover the constituency and ensuring that RSE teams at SSO institutions maintain a critical mass that will facilitate community development and knowledge transfer within the SSO. We expect it to be fairly easy to extend the SSO to other national laboratories and organizations as appropriate. Knowing the constituency the sponsor intends to support and the budget levels for the SSO will be important information to guide decisions about the ultimate institutional footprint of the SSO's RSE team.

We anticipate bringing additional partners into the SSO to support some of the activities of the project. We have already included representatives of the four ASCR user facilities (ALCF, ESnet, NERSC, and OLCF) as senior personnel at the seedling stage of the project. Our conception of the COLABS SSO is perhaps more like a user facility for software developers than a traditional ASCR R&D project, so their experience will be helpful. We expect to engage additional facility staff to support some of the SSO's services (e.g., CI pipelines, porting to high-end and emerging hardware, etc.).

For Training, two NSF supported projects that could be valuable partners are: the INTERSECT project [6], and the Trusted CI (cyberinfrastructure, in this case) project [20]. The senior leadership of the SSO is al-

ready engaged with both projects. We also expect to work closely with the US-RSE organization, which is the professional society for RSEs based in the U.S. It currently has working groups on “Diversity, Equity, and Inclusion”, "Education and Training", "Outreach", and “RSE Empowerment at National Labs”, all of which are quite relevant to the mission of the COLABS SSO and the interests of its workforce. And we intend to engage a small business with experience running educational and internship programs with strong IDEA components to ensure those programs are as effective as possible.

For the R&D effort, we anticipate bringing in a small number of researchers who focus on software engineering and cognitive and social science to ensure the quality of the research effort.

1.3.8 Potential relationships with other SSO concepts

If multiple SSOs are funded by ASCR, we intend to work cooperatively to maximize the value provided to the sponsor and client projects while minimizing redundancies. Whereas we focus on providing RSE capabilities to support stewardship needs on an individualized per-project basis, we expect that other SSOs may pursue other approaches. We are happy to treat other approaches as complementary to ours, in the spirit of giving multiple seedlings a chance to grow and try out different models for software sustainability. On the other hand, we believe that our RSE-based approach is very flexible, so, if desired (by the sponsors or by client projects), the COLABS SSO could provide similar services in many cases. At the same time, other SSOs may also plan training, R&D, or other activities which are similar to what we plan. In such cases, we would be happy to collaborate.

1.4 Expected outcomes and benefits

Our COLABS SSO concept offers a flexible and scalable approach to addressing the needs of scientific software projects. We believe that this will allow it to successfully address the software sustainability and stewardship needs of a substantial portion of the DOE scientific software base.

Tailored support for software projects. In any given constituency, different software projects will be at different places in their lifecycle, taking different approaches to development and having different goals and needs. Our RSE-based approach allows us to address the needs of client projects on an individualized basis, rather than a one-size-fits-all approach.

Separation of R&D from stewardship. Though the line between the two is not well-defined, our SSO concept focuses on stewardship, in support of decisions made by DOE program managers about the R&D activities they wish to support. We expect this to be the subject of ongoing discussions with the sponsors.

Building the software development workforce. We believe that RSEs will constitute a growing portion of the DOE scientific software workforce. Our approach will help build that community and make connections into educational institutions to build and fill a pipeline of future RSEs through education and experiential learning. This pipeline effort will also provide significant opportunities to help address issues of diversity, equity, and inclusivity in this workforce.

Organizational scalability. While our concept is anchored at ANL, LBNL, and ORNL, it is by no means limited to those sites. As long as we can maintain a critical mass of RSE staff at each institution, and as long as the institutions “understand” and can support the RSE role, the SSO’s institutional footprint can be scaled to be similar to the footprint of the constituency it is meant to serve.

Budgetary scalability. With some caveats, our SSO concept can scale to fit a variety of budget levels. The RSE-based services can be considered in terms of a number of full-time equivalents (FTEs) per project in the constituency, and the SSO will provide whatever services it can within the allocated support. The cost of Training and Outreach and R&D activities will scale less than linearly with the size of the constituency.

Scalability to additional sponsors. We believe that one of the key benefits of our approach is that the same structure and largely the same governance model can support multiple sponsors and constituencies.

Thus, for example, other offices within the DOE Office of Science (SC) might fund the SSO to provide sustainability services to their application software teams or user facility software suites. Sponsors could equally well include applied offices within DOE or other federal agencies that work closely with the DOE.

1.5 Risks and mitigation strategies

This kind of concerted effort on the sustainability and stewardship of scientific software is new, and there will undoubtedly be many challenges as we further develop and implement the COLABS SSO. Here, we briefly discuss some of the known risks and how we envision managing them.

Recruitment and retention of RSE staff. Research software engineers form the core of our approach to software sustainability. This role is often hard to staff at present. We plan to be very aggressive in recruiting, but focusing more on potential and interest than specific experience, then providing the needed training once they're on board (Sec. 1.3.1). In the near term, we anticipate that the wind-down of the ECP project will provide a pool of candidates who will be interested in RSE roles. In the longer term, we also expect to leverage our educational outreach and pipeline development efforts (Sec. 1.3.1). Retention poses a challenge because a well-trained RSE is also attractive to the industry. We will advocate for parity in status with scientific staff at the labs, clearer career paths, and more opportunities to contribute to interesting cutting-edge research projects to make the role more attractive.

Client project engagement. Our approach will work best if there is an active team within the client project to engage with us in a spirit of collaboration. SSO staff will receive training in the social aspects of their work to help guide the collaboration along the most effective lines. Periodic status review meetings between SSO management and project leadership will also touch on the relationship in addition to technical progress.

Managing client expectations. Staffing levels and funding levels will ultimately determine the level of effort that is available to support projects seeking the services of the SSO. We will need to be transparent about what clients can expect in this regard. In some situations, we might have to triage requests even at the Essential Services level (Sec. 1.3.2) to prevent RSE staff from being spread too thinly to be effective.

1.6 "Seedling" phase activities

During the short "seedling" phase which is the subject of the current proposal, our primary focus will be on working with ASCR and the prospective constituency to refine the COLABS SSO design, and as needed information and parameters become available, to begin working out details that will allow the actual SSO to be stood up as quickly a possible once funding becomes available. Key activities are outlined below.

Community building and data gathering. We plan to hold 1-2 online town halls or workshops to get community input to help us refine the list of services we'll offer and prioritize them. We may also use surveys to gather data about the constituent projects.

Discussions with other potential sponsors. We have designed the COLABS SSO to scale to accommodate additional sponsors with their own constituencies. We plan to engage with other potential sponsors regarding our approach, both to gather further input to refine our concept and to gauge interest in possible future expansion.

Discussions with ASCR and other seedling SSOs. So far, the development of our SSO concept has been little constrained by the sponsor. But going forward, we expect to engage with ASCR to better understand the "bounding box" that will apply to the operational SSOs. We hope to engage on topics such as the relationship between R&D and software sustainment, and the relationship that the program managers expect to have with the operations of the SSO. To move forward effectively, we will need an understanding of the budgetary expectations and details of the constituency envisioned. We expect that the other seedling SSOs will want to engage in similar discussions.

Coordination with other seedlings. As mentioned in Sec. 1.3.8, we expect that coordination with other seedling SSOs will be desirable to minimize redundancies. In some cases that may be in the form of collaboration (e.g., on training), while in others efforts would be made to ensure that offerings are complementary.

Refining the governance model, management approach, and budgetary model. We have outlined our initial ideas about the governance of SSO services (Sec. 1.3.6) and management (Sec. 1.3.5). These need to be refined so that they are ready to put into operation. Further development of the governance model will be informed by the discussions with the community, with ASCR, and other potential sponsors. A clearer picture of the services the SSO will offer and terms of engagement with client projects will allow revision of the WBS and the development of a more detailed management plan. We also expect to line up members for the advisory committee during the seedling phase. Finally, we will develop a budget model which will guide the distribution of resources across the various services and activities of the SSO and ideas of what is feasible at different levels of funding.

Team and staffing. A major activity during the seedling phase will be expanding the SSO team in terms of institutions and individuals. Once we have a clearer picture of the nature of the constituency (including its institutional footprint), and the (approximate) budget levels expected for the operational SSO, we can start making decisions about institutional partners for the RSE-based services and other needs. In other areas of the project, we have more specific ideas about who we want to engage, as discussed in Sec. 1.3.7, and we can start bringing in those people and organizations. We will also work with ASCR, and the management of member institutions and the ECP to develop a plan that will allow us to reach out to ECP team members about follow-on opportunities in the SSO without disrupting ECP projects in their wind-down phase.

1.7 Summary

We have outlined a vision for a software sustainability organization which will address major software sustainability needs of Office of Advanced Scientific Computing Research software projects and potentially those of other sponsors. At the heart of our concept are research software engineers who will provide a wide range of services to client projects, tailored to their needs. The SSO will also provide extensive training in software engineering practices, outreach to elevate software as a first-class scientific product and promote the recognition of its developers, as well as a modest R&D effort to improve the SSO's and the community's understanding of scientific software and how to do a better job of developing and sustaining it. And it will be active in the development and retention of RSEs through educational and internship programs at the university level and efforts within the SSO itself.

Our SSO approach makes the distinction between R&D and stewardship, and it will help build the software development workforce that the modern scientific enterprise increasingly requires. The design is scalable to embrace additional organizations joining the SSO as providers of RSE services, scalable with the budget, and scalable to support additional sponsors with different constituencies.

During the seedling phase of the SSO development, covered by this proposal, we will reach out to the community to introduce our approach and gather input to further refine our plans, work with ASCR to better understand their expectations of how the SSO will operate and to flesh out key concepts in the design, including the governance model, management approach, and budgetary model. Finally, once appropriate details are known, we can finalize the team for the operational SSO and begin recruiting staff.

Appendix 3. References cited

[1] Society of Research Software Engineering, https://society-rse.org/about/.

[2] IDEAS Productivity, Events, https://ideas-productivity.org/events/, 2022, accessed 2022-12-17.

[3] IDEAS Productivity, Responding to the Software Crisis in DOE Scientific Computing, http://ideas-productivity.org/resources/workshops/2015-09-software-crisis/, 2015, accessed 2022-11-28.

[4] Michael A. Heroux and Gabrielle Allen, Productive and Sustainable Computational Science & Engineering Software Productivity and Sustainability (CSESSP) Challenges Workshop Report, Technical report, Networking and Information Technology Research and Development (NITRD) Program, Arlington, VA, 2016, available at https://www.nitrd.gov/PUBS/CSESSPWorkshopReport.pdf, accessed 2022-11-28.

[5] BSSw Fellowship Program, https://bssw.io/fellowship, 2022, accessed 2022-12-01.

[6] INTERSECT – Research Software Engineering Training, https://intersect-training.org/, accessed 2022-11-29.

[7] IDEAS Productivity Project, The Better Scientific Software (BSSw) Tutorial, https://bssw-tutorial.github.io/, 2022, accessed 2022-11-28.

[8] IDEAS Productivity, Best Practices for HPC Software Developers (Webinars), https://ideas-productivity.org/events/hpc-best-practices-webinars/, 2022, accessed 2022-11-28.

[9] Better Scientific Software (BSSw), https://bssw.io/, 2022, accessed 2022-11-28.

[10] Argonne National Laboratory, Leadership Institute, https://leadershipinstitute.anl.gov/, accessed 2022-12-17.

[11] Michael A. Heroux, Elsa Gonsiorowski, Rinku Gupta, Reed Milewicz, J. David Moulton, Gregory R. Watson, Jim Willenbring, Richard J. Zamora, and Elaine M. Raybourn, Lightweight Software Process Improvement Using Productivity and Sustainability Improvement Planning (PSIP), in Guido Juckeland and Sunita Chandrasekaran, editors, *Tools and Techniques for High Performance Computing*, pages 98–110, Cham, 2020, Springer International Publishing.

[12] Argonne Leadership Computing Facility, https://www.alcf.anl.gov/, accessed 2022-12-17.

[13] ESnet: Energy Science Network, https://www.es.net/, accessed 2022-12-17.

[14] National Energy Research Scientific Computing Center, https://www.nersc.gov/, accessed 2022-12-17.

[15] Oak Ridge Leadership Computing Facility, https://www.olcf.ornl.gov/, accessed 2022-12-17.

[16] David E. Bernholdt, John Cary, Michael Heroux, and Lois Curfman McInnes, The Science of Scientific-Software Development and Use, brochure from the basic research needs workshop, https://doi.org/10.2172/1846008, 2022.

[17] Software Sustainability Institute, https://www.software.ac.uk/, accessed 2022-12-17.

[18] Carnegie Mellon University, Software Engineering Institute, https://sei.cmu.edu/, 2022, accessed 2022-12-17.

[19] The United States Research Software Engineer Association, https://us-rse.org/, 2022, accessed 2022-11-29.

[20] Trusted CI, the NSF Cybersecurity Center of Excellence, https://www.trustedci.org/, 2022, accessed 2022-11-29.

[21] Argonne National Laboratory, Diversity and Inclusion, https://www.anl.gov/diversity-and-inclusion, accessed 2022-12-18.

[22] Berkeley Lab, IDEA: Inclusion Diversity Equity Accountability, https://diversity.lbl.gov/, accessed 2022-12-18.

[23] Oak Ridge National Laboratory, Diversity, Equity, and Inclusion, https://diversity.ornl.gov/, accessed 2022-12-18.

[24] AnitaB.org, Grace Hopper Celebration, https://ghc.anitab.org/, 2022, accessed 2022-12-18.

[25] CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing Conference, https://tapiaconference.cmd-it.org/, 2022, accessed 2022-12-18.

[26] National Laboratory Women of Color Community Forum, https://blogs.anl.gov/nlwoc/, accessed 2022-12-18.

[27] Why Women Don't Apply for Jobs Unless They're 100% Qualified, https://hbr.org/2014/08/why-women-dont-apply-for-jobs-unless-theyre-100-qualified, 2014, accessed 2022-12-18.

[28] 6 Common Hiring Practices With Unintentional Bias, https://www.entrepreneur.com/living/6-common-hiring-practices-with-unintentional-bias/317858, 2018, accessed 2022-12-18.

[29] Definition of imposter syndrome, https://www.merriam-webster.com/dictionary/impostor%20syndrome, accessed 2022-12-12.

[30] Samyukta Mullangi and Reshma Jagsi, Imposter Syndrome: Treat the Cause, Not the Symptom, *JAMA* **322**, 403 (2019).

[31] Maddie Breeze, Imposter Syndrome as a Public Feeling, in Yvette Taylor and Kinneret Lahad, editors, *Feeling Academic in the Neoliberal University: Feminist Flights, Fights and Failures*, pages 191–219, Springer International Publishing, Cham, 2018.

[32] Anshu Dubey, Katie Antypas, Murali K Ganapathy, Lynn B Reid, Katherine Riley, Dan Sheeler, Andrew Siegel, and Klaus Weide, Extensible component-based architecture for FLASH, a massively parallel, multiphysics simulation code, *Parallel Computing* **35**, 512 (2009).

[33] Center for Scientific Collaboration and Community Engagement, Scientific Community Engagement Fundamentals, https://www.cscce.org/trainings/cef/, accessed 2022-12-10.

[34] Dale Carnegie, *How to Win Friends & Influence People*, Arrow (Random), 2012.

[35] Allan R. Cohen and David L. Bradford, *Influence Without Authority*, Wiley, 3rd edition, 2017.

[36] Karen Catlin and Sally McGraw, *Better Allies: Everyday Actions to Create Inclusive, Engaging Workplaces*, Better Allies Press, 2nd edition, 2021.

[37] Everett M. Rogers, *Diffusion of Innovations*, Free Press, 5th edition, 2003.

[38] Stanley McChrystal, Tantum Collins, David Silverman, and Chris Fussell, *Team of Teams: New Rules of Engagement for a Complex World*, Portfolio, 2015.

[39] Sustainable Horizons Institute, Sustainable Research Pathways, https://shinstitute.org/sustainable-research-pathways-2022/, 2022, accessed 2022-12-18.

[40] Argonne National Laboratory, Science Careers in Search of Women, https://www.anl.gov/science-careers-in-search-of-women, accessed 2022-12-10.