

2022 CASC Top Priority Areas: Support Team Science and End-to-End Workflows

Summary Statement

Institutions should provide sustained support for resources that serve the entire institution, develop methods and governance structures to coordinate the activities of core research and campus-based Research Computing and Data (RCD) resources, recognize RCD professionals' role, including training them to stay at the leading edge of their fields, and develop a culture focused on the use of teams. Research sponsors should provide funding opportunities for campuses to address emerging RCD infrastructure needs, incentivize participation of RCD professionals in funded research projects, continue to provide funds to develop tools designed to lower barriers to Team Science, and promote and support the development of networks of relevant RCD professionals. CASC should highlight examples of how member institutions are addressing challenges related to the support of Team Science at its annual meetings and through its publications and website, and should explore ways to partner with other organizations who also have a stake in advancing support for Team Science.

Why it Matters

The evidence for the positive role that Team Science has had in advancing knowledge in all areas of scientific inquiry is overwhelmingly strong¹. The National Institutes of Health (NIH) defines Team Science as “a collaborative effort to address a scientific challenge that leverages the strengths and expertise of professionals, oftentimes trained in different fields.” Our intention in this position paper is to highlight why Team Science is directly relevant to the work that our members are engaged in as leaders and practitioners building and maintaining advanced research computing and data infrastructures.

To a large extent, Team Science matters to our members because it has become the prevalent approach to conducting research. The shift in research away from an individual investigator toward teams of investigators working collaboratively is now the norm and our efforts to better accommodate this new norm have intensified in recent years and have helped to lower the barriers to effective Team Science.

More than ever before, the RCD facilities we operate are a nexus for large and small multidisciplinary collaborations. Where entire experiments were once conducted from beginning to end by a single graduate student on campus or remote High Performance Computing (HPC) systems, it is now common for the HPC system to serve as a part of a larger workflow, which might include remote instruments, such as DNA sequencers, microscopes, or a repository of remote data, such as multispectral images from satellites. The facilities that we operate may in turn generate data that serve as inputs to downstream decision tools managed by external entities and used by domain experts with varying levels of computational experience and expertise.

Better integrating research computing and data systems into the broader research fabric of our campuses and of the nation is essential to sustaining the advances made possible by Team Science. Advances in RCD that work to further support Team Science will help to catalyze a virtuous cycle of development that encourages researchers to reimagine what is possible, which will in turn challenge our members to develop new ways to lower barriers to accessing and integrating advanced research computing and data storage resources into large and small scale workflows. We see no near-term end to this virtuous development cycle and recognize that the more we do to promote it, the further we will advance Team Science and by extension, research discoveries.

¹ [1] Wuchty, S., Jones, B.F. and Uzzi, B. 2007. The increasing dominance of teams in production of knowledge. *Science*. 316, 5827 (May 2007), 1036–1039. DOI:<https://doi.org/10.1126/science.1136099>.

In the best of worlds, end-to-end workflows bring together IT professionals, software engineers, molecular biologists, computational scientists, and mathematicians as a team with a common focus to find ways to leverage advanced detectors and instruments, fast networks, advanced computational platforms, and improved algorithms to produce entirely new insights and discoveries. These diverse teams working on every imaginable type of instrument are generating results that are fundamentally transforming the way we think about the world in fields like agriculture, medicine, and the environment. Well-run RCD resources are an essential component of Team Science, which is why this position is among our member's top concerns.

Current Challenges

While progress has been made, the challenges our members face persist. There is still a need to develop solutions that lower barriers to integrating research computing and data resources more broadly within our campus and national research infrastructures, such that we are better equipped to support Team science. These challenges fall under a broad set of areas, including cultural, technical, financial, operational, legal, regulatory, and strategic considerations.

Cultural and generational challenges start from the lack of common terminology to describe these challenges, which contribute to a lack of common awareness and understanding and to differences in prioritization among different portions of the community. These challenges can have negative effects on timeliness of engagement and effectiveness of approaches to finding solutions to these problems. Technical considerations such as lack of common tools and/or difficulty in porting tools between environments aggravate such cultural challenges, leading to non-uniformity of user and developer tools. Security considerations often reduce opportunities to collaborate and develop common frameworks.

Growing legal and regulatory frameworks have accelerated the need for operational expertise in advancing cyberinfrastructure in Team Sciences. For research teams to navigate successfully through the complex issues of intellectual property, privacy and compliance, expertise is needed in the areas of researcher training, project management, and effective communications to stakeholders. There is no clear strategy that prioritizes RCD efforts, especially financially. This situation is aggravated by a general lack of planning and investment in research computing infrastructure, both on the part of funding agencies and the research community, as well as by a lack of reliable funding sources for long-term Team Science oriented RCD efforts.

Technical debt has also been a long-standing problem in the RCD arena that aggravates the assembly of effective tooling in team-based science. Issues include the lack of common tool sets in many areas, difficulty in porting legacy or familiar tools to new RCD ecosystems, and non-uniformity in skills for software development and use as well as for data management.

Financial and operational issues in Team Science include lack of funding or failure to fund development needed to support building and maintaining teams, lack of onboarding and training programs in certain areas causing a high threshold for learning for new participants, lack of coordination and communication among community participants and between these participants and the funding agencies, and a frequent lack of project management expertise in areas in which Team Science approaches are newly forming.

Beyond these issues, Team Science topics are subject to the same needs for cybersecurity, legal and regulatory compliance, privacy, intellectual property protection, and strategic prioritization of funding areas and approaches that are discussed more generally elsewhere in other CASC position papers.

Vision of Success

Our vision of success for Team Science is supported by three main foundational pillars: 1) a well supported RCD infrastructure geared to supporting Team Science; 2) external funding opportunities designed to incentivize new approaches to conducting research and discovery and fill in gaps where existing campus-based and national RCD infrastructures fall short; and 3) a skilled workforce capable of addressing cross disciplinary research and development challenges. These concepts are described in greater detail below.

Well Supported Campus RCD Infrastructure

Breaking free from the limitations of RCD infrastructure designed primarily for single investigators and relatively isolated workflows will require a combination of technical and culture changes. Success supporting Team Science is predicated on our campus leaders prioritizing support for shared RCD resources. Similar to libraries, campus-based RCD resources and services must be viewed as essential for there to be broad, multidisciplinary collaborations and seamless access to public and private remote RCD resources. At a minimum, this means that campus leaders will sustain investments in centralized and well-coordinated RCD resources.

Building on this RCD foundation, campuses will also invest in the coordination of activities and infrastructure among their core shared research facilities. These investments will be made manifest in sustained support for hardware and software resources and governance that will help to coordinate operations, set priorities, and attend to strategic opportunities. In this future state, we will have overcome, or at least minimized, most of the technical obstacles that make it difficult to support diverse assemblages of researchers and physical resources that are essential for Team Science.

For example, the task of tracking resource utilization back to single investigators, while important, currently involves a large amount of effort for both the investigators and our RCD professionals. The issue of tracking resource utilization is particularly problematic with Team Science because a team will likely consist of members who may come from any number of faculty labs, different colleges, campus core research facilities, and even collaborators outside of the primary investigator's home institution.

We envision a future state where core research facilities (AKA, Shared Research Facilities) and campus-based and national RCD resources perform as interoperable components that can be assembled to create small to large workflows in much the same way as methods can be assembled and used for different purposes as is done with Object Oriented Programming.

Finally, our vision of a campus that is geared to support Team Science is one that promotes a culture where faculty are encouraged to engage with RCD professionals during the planning stages of a project and collaborate on how best to coordinate, assemble, and deploy RCD resources in support of the project's research goals.

Sponsor Support to Address Emerging Challenges and Fill Gaps

Public and private research sponsors play a critical role in the advancement of our nation's R&D agenda. We envision a future state where sponsors play a greater role in helping to support the RCD resources that are essential to the advancement of Team Science. For example, future solicitations will explicitly emphasize the need to engage with campus-based RCD professionals during the proposal planning stage. Proposed research plans that require a higher level of workflow development and/or coordination will be encouraged to include funds for RCD professionals, so that these individuals may remain engaged with the project.

Funding for physical RCD infrastructure, software, and staff will be available to address new and emerging areas where campuses may not yet have provided support. For example, funding will be more broadly available to support the development of tools to facilitate Team Science, both generally and for specific domains of science. In recognition of the critical role that RCD professionals play in Team Science, funding will be available to promote information sharing related to Team Science.

These networks of RCD professionals spanning our nation's universities are essential elements of sponsored research activities in general and Team Science challenges in particular. Hence, it is in the interest of our research sponsors to include support that will help to nurture, maintain and grow these communities so that they are ready to address current and emerging challenges.

An RCD Workforce Skilled in Team Science

The critical role that RCD staff play in the support of Team Science cannot be overstated and is described in greater detail in the "Staffing Challenges" section of this document. A vision of success for Team Science would not be complete without at least some mention of the importance of a diverse and well trained RCD workforce with specific expertise in this area. RCD staff play a unique role in teams whose constituents are traditionally composed of faculty, post-doctoral fellows, and graduate students.

Just as faculty and students bring unique attributes to the team, RCD staff possess technical knowledge and skills developed over a greater period of time than most undergraduate, graduate, and postdoctoral students generally have time to accumulate. These include experience with Team Science concepts and tools needed to implement them. The RCD professional has a well developed view of what resources are available and how these can be brought to bear on end-to-end workflows that may be required for the success of each team, including DEI considerations as discussed in the previous .

A vision of success therefore includes having RCD staff available on campuses and at institutions with the skills needed to address the dedicated demands of Team Science. Funding from sponsors will help to support these professionals, so that more professionals may be hired to scale with demand. Not only are these professionals diverse in terms of the skill that they possess to address what the Campus Research Computing Consortium (CaRCC) calls research "facings"², but they will also more equitably represent genders and ethnicities.

Working Toward the Vision

The following links lead to examples of how members of the RCD community are supporting one or more of the pillars described above. These examples highlight what is being done and offer a direction that could be taken that will help to contribute to the broader vision of support for Team Science.

Recommendations and Next Steps

While there is no single "silver bullet" that will magically remove the barrier and challenges that we face in our efforts to support Team Science on our campuses, there are tangible steps that can be taken to accelerate progress toward the vision of success described above. This section lays out in more concrete terms what can help to advance this vision by applying best practices within our respective institutions, promoting positions that emphasize what our sponsor can do to help support Team Science, and taking specific actions within CASC and in partnership with other similarly aligned organizations.

² https://carcc.org/about/#define_rcd

Best Practices for Member Institutions

1. **Provide sustained support for local RCD resources that serve the entire institution.** Well maintained local RCD resources that are available to researchers from the moment that they begin their work at your institution are essential to the advancement of Team Science. Without such core resources researchers will spend an inordinate amount of time and resources developing fundamental RCD components that should be in place from the very beginning of their tenure.
2. **Develop means to coordinate the activities of core research facilities (AKA, shared research facilities) and campus-based RCD facilities.** Team Science depends on workflows that typically span the administration boundaries of multiple independent research cores. Institutions should develop governance and technical approaches that minimize the “friction” that often occurs at the intersection of these different campus resource cores. The friction may be the result of technical, financial, or cultural differences among the participating entities.
3. **Recognize the vital role that RCD professionals play at your institution and train them to stay at the leading edge of their field.** RCD staff must understand the broader RCD ecosystem including data generating instruments on campus and remote RCD resources, such as publicly funded national research computing grids and cloud capabilities. Part of the institutional investment in RCD should focus on recruiting and retaining underrepresented professionals with a broad set of technical expertise and “soft skills” where relevant. Consider developing a project management group to help manage and coordinate projects that have end-to-end workflows. Offer professional training opportunities that will help your staff develop and strengthen skills related to mediation, negotiation, and conflict resolution. Such skills will be particularly helpful when working with large multidisciplinary research teams.
4. **Develop a culture that designs research projects around the notion of team.** Encourage faculty developing grant proposals to engage with RCD facilitators at the very beginning of the grant writing process. This can be accomplished by taking steps to increase awareness of core campus resources that can be used to advance Team Science. Other incentives can be applied to promote the use of core resources, including providing start-up funds or “credits” that can be used to retain dedicated RCD staff who are well-positioned to implement workflows. Similarly, institutional seed funding can be used that gives people resources that will also help to build institutional capacity to support other similar projects.

Positions to Research Sponsors

1. **Provide funding opportunities for campuses to address emerging RCD infrastructure needs.** Leading-edge research activities often push the envelope of what resources can be provided by campus RCD organizations. Providing funding opportunities for such physical resources gives institutions a way to fill these gaps. Where appropriate, specialized resources can also be shared via national RCD organizations or at a minimum, lessons learned integrating these leading edge resources can be shared broadly with the RCD community.
2. **Incentivize participation of RCD professionals in funded research projects.** RCD professionals play a unique role in Team Science that is not filled by tenure-track faculty, post doctoral researchers, and graduate students. Not only do they have advanced knowledge of the RCD ecosystem on campus and beyond, but they also play an important role in mentoring students and developing the workforce that is the lifeblood of the private sector.

3. **Continue to provide funds for the development of tools designed to lower barriers to Team Science.** Large end-to-end workflows that are often an important component of Team Science still face a number of technical challenges. Funding to address these challenges would incentivize innovative approaches and that could be more broadly applied to a larger network of institutions and application areas. These technical challenges compound the problems our community faces related to the generation of reproducible workflows.
4. **Promote and support the development of networks of RCD professionals.** The role of RCD professionals is critical to the advancement of Team Science. Without their support, our nation's research institutions will not be able to support grand challenge research. Therefore, it is in the interest of our research sponsors to provide the support that will help nurture, maintain, and grow these communities so that our research institutions are well prepared to address current and emerging challenges.

Actions by CASC

1. **Highlight examples of how member institutions are addressing challenges related to the support of Team Science at CASC annual meetings.** The example should get into the details of the challenges that institutions face, not just the hardware resources that we deploy, but the relationships and trust we must develop over our entire institution. This is a case where members are interested in learning "how the sausage was made." Discussions could also include information about what campuses are doing to provide training opportunities to our members in areas such as project management, conflict resolution, and other areas that are needed to help advance Team Science.
2. **Explore ways to partner with organizations who also have a stake in advancing RCD support for Team Science.** Connect with aligned organizations to discuss what a partnership would look like. Work with partner organizations to define, update, and promote the importance of RCD professionals and the concept of "Workflow engineering" as a service.