Fostering Collaboration Among Organizations in the Research Computing and Data Ecosystem

Sharon Broude Geva*
Office of Research
University of Michigan
Ann Arbor, MI, USA
sgeva@umich.edu

Dana Brunson
Internet2
Stillwater, OK, USA
dbrunson@internet2.edu

Thomas Cheatham III
Dept. of Medicinal Chemistry
University of Utah
Salt Lake City, Utah, USA
tec3@utah.edu

James Deaton
Great Plains Network
Columbia, MO, USA
jed@greatplains.net

James Griffioen
Center for Computational Sciences
University of Kentucky
Lexington, KY, USA
griff@uky.edu

Curt Hillegas
Research Computing
Princeton University
Princeton, NJ, USA
curt@princeton.edu

Douglas Jennewein
Research Technology
Arizona State University
Tempe, AZ, USA
douglas.jennewein@asu.edu

Gail Krovitz
Internet2
Denver, CO, USA
gkrovitz@internet2.edu

Tobin Magle
Research Data Services
University of Wisconsin
Madison, WI, USA
tobin.magle@wisc.edu

Patrick Schmitz
Semper Cogito Consulting
Oakland, CA, USA
patrick@sempercogito.com

Karen Tomko
Ohio Supercomputing Center
Columbus, OH, USA
ktomko@osc.edu

James Wilgenbusch
Research Computing
University of Minnesota
Minneapolis, MN, USA
jwilgenb@umn.edu

ABSTRACT

The widespread application and success of computational and data intensive research approaches in every discipline represented on our campuses has resulted in a rapid proliferation of organizations, technologies, and professions affiliated in different ways with the support and advancement of activities related to research computing and data (RCD). While most agree that this growth is helping to advance numerous disciplines, the proliferation of organizations seeking to support, promote, and advance RCD has led to some challenges. Specifically, a lack of understanding and consensus concerning which organizations should be considered a part of research computing hampers our ability to encourage collaborations among its complementary constituents, leads to unneeded and redundant activities, and makes it difficult to identify strategic priorities and address gaps where specific needs are not being met to advance various disciplinary activities. In this paper we introduce the ecosystem metaphor to help characterize the rapidly changing relationships between the growing set of organizations that in some way support and enable activities related to RCD. The ecosystem concept lends itself well to describing the many entities related to RCD because it emphasizes the larger system over its individual component parts and highlights their interdependence, while explicitly expecting their change over time. Our work to characterize the current RCD ecosystem, while imperfect, will serve as a foundation and framework for the development of a more complete view of the ever-changing RCD ecosystem. A more complete view of the RCD ecosystem will in turn help to advance the broad goals of its members by helping to foster and accelerate new and meaningful collaborations among them.

CCS CONCEPTS

• **Social and professional topics** → *User characteristics*; • **Human-centered computing** → *Human computer interaction (HCI)*;

KEYWORDS

research computing, research data management, research IT, ecosystem; community, cyberinfrastructure, people

ACM Reference Format:
Sharon Broude Geva, Dana Brunson, Thomas Cheatham III, James Deaton, James Griffioen, Curt Hillegas, Douglas Jennewein, Gail Krovitz, Tobin Magle, Patrick Schmitz, Karen Tomko, and James Wilgenbusch. 2020. Fostering Collaboration Among Organizations in the Research Computing and Data

*Authors are listed alphabetically.

PEARC '20, July 26–30, 2020, Portland, OR

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Ecosystem. In *PEARC '20: Practice and Experience in Advanced Research Computing, July 26–30, 2020, Portland, OR.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Over the last decade, the number of researchers who require advanced support for computation and data to carry out their research has grown dramatically. This growth is driven by the widespread application of computational and data intensive research approaches beyond the desktop across nearly every discipline represented on campuses today. This is amplified by the explosive growth of, and easy access to, massive computational capabilities and shared data sets both in the public and commercial realm.

The ability to leverage computation and data to advance research – which we will hereafter refer to as *research computing and data (RCD)* – is drastically changing the way researchers across all disciplines carry out their research, allowing them, for example, to computationally and virtually perform tasks and experiments previously only done in an experimental lab, to collect unprecedented amounts of detailed data via sensors and mobile devices, to share (individually insignificant) information that can be combined and aggregated into large statistically meaningful data sets that drive discoveries, or to automate the research discovery process via machine and deep learning. This transformation is due in part to a combination of changes that span technology (e.g., hardware and software improvements), social engineering (e.g., an emphasis on cross-disciplinary research), political forces (e.g, the development of national research facilities), and economic factors (e.g., the emergence of commercial cloud services). However, the same technological, social, political, and economic advances that have enabled RCD to be such a transformational force are also creating new challenges and complexities for researchers and the organizations that support them. These challenges range from sustainability to workforce development to training and access to resources. Despite the need and desire to work together, perhaps due to lack of communication and understanding of the entire ecosystem, these organizations and communities often feel they operate in unrelated areas and do not fully understand or appreciate the value of working together. In many cases they run in different (social and professional) circles – think conferences – and thus are unaware of one another and/or each other's approaches to these problems. Often they do not play well together for structural or political reasons, or lack adequate (economic) ways to compensate one another, or to exchange services.

Organizations and communities that support some aspect of RCD within and between campuses already exist in the areas of high performance computing system administration, (big) data storage, network connectivity, HPC education, test beds and experimentation, resource sharing, research libraries/standards, professional development, best practices/guidelines, and diversity and inclusion. All have played a role in shepherding some aspect of the complex RCD system that is emerging today. For the most part, these organizations have evolved independently, serving their particular niche community. However, over time these communities have matured and are beginning to develop into an interdependent collection of organizations that overlap in parts of their mission and activities and that through improved collaboration have enormous potential to collectively transform and support research computing and data science.

To fully capitalize on this potential will require that these communities organize into a working *Ecosystem*, fostering collaboration, coordination, and awareness, collectively working to advance the goals of the ecosystem as a whole while removing unnecessary redundancy to enable a broader scope of coverage of the ecosystem needs. Unfortunately, a lack of complete understanding and consensus concerning which organizations should be considered a part of RDC hampers our ability to encourage collaborations among its complementary constituents. This leads to unneeded and redundant activities, and makes it difficult to identify strategic priorities and address gaps where specific needs are not being met to advance various disciplinary activities

The purpose of this paper is threefold. First, we will describe our approach to both identify the organizations that currently comprise the Research Computing and Data (RCD) Ecosystem and to understand the relationship among them. Second, we will characterize the current relationships among these organizations to establish a baseline by which changes among the organizations that make up the RCD ecosystem over time can be better understood. Third, we outline how we plan to enhance our awareness, communications between, and understanding of the organizations that comprise RCD and call for participation.

The remainder of the paper is organized as follows. Section 2 provides some examples of ways in which collaboration between organizations could benefit and provide synergy between researchers, institutions, and funding agencies. Section 3 describes the approach and steps we have taken to identify members of the RCD ecosystem and understand their relationships. Section 4 presents preliminary data that has been collected about the capabilities and resources available across the RCD community, and Section 5 concludes the paper with our plans and next steps.

2 POTENTIAL VALUE OF A FUNCTIONAL RCD ECOSYSTEM

To illustrate the potential value of a functional RCD ecosystem, we present three example anecdotes that highlight the ways in which collaboration among organizations, limiting redundancy, and identifying gaps and unmet needs can benefit research at all levels. The first example considers the problem of helping new and established organizations find resources needed to support their computation and data needs. The second example illustrates the advantages that can occur from collaboration and exchange of information between organizations, in particular, as it relates to storing, accessing, sharing, and archiving data – a growing challenge for researchers. The third example shows how a complete understanding of the ecosystem as a whole can help university administrators better understand the positioning of their center/institute in the national RCD ecosystem, and can highlight gaps and priority investment areas for funding agencies.

2.1 Example 1: How a clear view of the RCD Ecosystem can help advance the broad goals of new organizations with RCD needs

Advances in most disciplines now require broad collaborations that cut across traditional disciplinary boundaries. Notably this includes leading-edge RCD infrastructure and services (or people) that can effectively collect, integrate, analyze, manage, and share complex data sets and results. As new disciplines push their traditional boundaries to insight and discovery, there exist outstanding opportunities to leverage lessons learned. We are increasingly dependent on complex and inter-dependent workflows, which in turn rely on sufficient RCD infrastructure and services, which are constantly evolving. Efforts are accelerated when those seeking to develop new and domain specific workflows and tools have a good picture of the rich and available set of organizations, infrastructure, and professionals that make up the RCD ecosystem and are already engaged in similar activities. Likewise, new domains of research can more effectively broaden their impacts if they are given a clearer view of the resources already available to share and promote new innovations of this kind. Finally, members representing new disciplines are able to more effectively broaden the participation of underrepresented groups when they can locate and tap into like organizations, which are affiliated with the RCD ecosystem. In response to cross disciplinary needs, a number of organizations have emerged, both old and new, to support the community with leading practices or advice on how to support RCD on campuses and beyond. Three relevant organizations that fill this function are CASC (Coalition for Advanced Scientific Computation), CC (Campus Champions), and CaRCC (Campus Research Computing Consortium). CASC has long had the mission of supporting leadership efforts in RCD and serving as a forum for sharing leading practices, as well as advocating for advanced research computing to funding agencies and campus stakeholders. Campus Champions emerged out of XSEDE (Extreme Science and Engineering and Discovery Environment) as a community to help guide people to XSEDE resources at NSF Supercomputing Centers but quickly broadened into a community sharing leading practices on campuses. Not only was this community a place to ask questions and get good answers on RCD questions, but it quickly became a community of peers with an identity. E.g., "Wow, I am providing this support on my campus, but I did not realize there were others doing the same thing; let's work together and help each other!". As the ACI-REF (NSF OAC- 1341935) project progressed, it was realized that not only do we need a community of researcher-facing professionals, but that other"facing" (i.e., roles) would benefit from community organizing as well, including systems-facing, data-facing, and stakeholder-facing, among others. The Campus Research Computing Consortium (CaRCC; https://carcc.org) emerged in order to advocate and ultimately develop the RCD profession with job families, capability models describing levels of uptake and competency across the RCD facings, and to help define stakeholders and value propositions. CaRCC is seen as a collaborative boundary spanner in the community and organized the first ecosystem workshop in April, 2019.

2.2 Example 2: How a clear view of the RCD Ecosystem can help researchers find solutions to their pressing research needs, such as effective and compliant data management

The days of manageably small data sets held privately by researchers with no need to share the data, and few – or relatively lax – privacy/security requirements, have largely become a thing of the past. Today's researchers are faced with complex data management challenges that span all stages of the data lifecycle. To make things worse, the expertise needed to manage research data through its entire lifecycle is spread across many organizations ranging from IT/HPC groups, to data carpentry groups, to library and information sciences groups, to national standards groups and repositories, to institutional review boards, to compliance groups, and, increasingly, groups focused on reproducible research with publications that contain data and artifacts. Because many of these organizations have grown and existed independently for many years, the connections between organizations are often weak to non-existent, leaving institutions and their researchers to "roll their own" solutions to today's data management problems. With increasing pressure from funding agencies on researchers and institutions to develop robust, scalable, long-term solutions to manage research data, it is increasingly important to develop connections between and across these organizations in an attempt to address ubiquitous research data management issues facing all research institutions. These issues include, but are not limited to, sharing, locating, accessing, searching, processing, reproducing, and archiving research data.

2.3 Example 3: How a clear view of the RCD ecosystem can help universities and foundations prioritize investments

Universities and foundations can begin to make strategic decisions about how to prioritize investments to accelerate research when armed with a coherent and global understanding of the organizations, infrastructure, and professionals that contribute to the advancement of compute and data intensive research and the interrelationship among these entities. Put another way, to know what you have is the first step to knowing what is missing. With the relatively recent and rapid increase in domains of research with compute and data intensive needs, most practitioners view the RCD ecosystem from their particular research perspective. University administrators are constantly asked to make long-lasting strategic investments in campus infrastructure. When a myopic, single-discipline view of the RCD ecosystem is presented, universities run the risk of making investments that only narrowly address challenges related to RCD and miss opportunities to collaborate with existing organizations already serving similar or complementary functions. Failing to engage more broadly with members of the RCD ecosystem not only negatively affects campus-based research communities, but limits the impacts that come from broader perspectives and contributions that can be made by a more diverse set of practitioners. Research foundations are in a position to incentivize the development of communities of practice that transcend single disciplinary activities. For example, NSF's Infrastructure

Capacity for Biology (ICB) program, "supports the development, expansion, or improvement of infrastructure that will enable fundamental research within the biological sciences." While the focus of this program is on biologically relevant applications, the expectation is that innovations under this program will be both leveraged and shared by other science and education communities. To do this effectively, a clear vision of the broad RCD ecosystem is needed.

3 APPROACH

CaRCC aims to advance the frontiers of research at academic institutions by supporting on-campus awareness and facilitation services related to computation for researchers, including inter-institutional resource and knowledge sharing among research computing professionals, and continuous innovation in research computing capabilities. It is committed to supporting the sustainability of campus efforts through professional and career development for the individuals who enable and collaborate with researchers to better utilize large-scale, advanced computing resources. CaRCC is further dedicated to extending and enhancing the reach and impact of campus and national research computing infrastructure on research conducted at the campus level (including multi-institution collaborations) by exploring and developing effective strategies and leading practices that campuses may use to empower their researchers to become more effective users of RCD.

As part of these efforts, CaRCC convened members of the larger community of communities supporting RCD to better understand current limitations to advancing domains of research that are highly dependent on RCD resources. The clear consensus among those gathered was that more could be done to promote a broad awareness of the different groups helping to advance RCD and lower barriers to their collaboration and interaction. This led to the formation of a volunteer working group in the fall of 2018 to design a workshop to bring together an initial set of RCD organizations to connect community members and organizations to leverage one another and to explore and advance the ever-changing ecosystem of RCD. An aim was to identify gaps not well serviced in the current ecosystem. The working group developed a set of guiding principles to create a convivial, inclusive, and open environment, which entailed the follow key elements:

- A commitment to respecting independent identities.
- A goal of accomplishing together what we can't do separately.
- An overarching shared value to make Research Computing and Data a true profession.
- A commitment to provide enabling RCD resources and services for research and researchers within and across campuses.

The workshop was co-located with the Coalition for Networked Information in St. Louis in April 2019. Session goals included:

- Identify key organizations and initiatives in the research computing and data ecosystem.
- Appreciate unique areas of expertise and points of overlap among the organizations and initiatives.
- Develop a shared visualization representing the ecosystem.
- Generate potential operating principles under which the organizations and initiatives can maintain their independence, while also advancing their points of interdependence.
- Ensure mechanisms for communication, coordination, and collective impact.
- Specify next step action implications.

Organizations represented at the workshop:

- Association of Research Libraries (ARL)
- Big Data Hubs
- Campus Champions (CC)
- Campus Research Computing Consortium (CaRCC)
- Carpentries
- Coalition for Academic Scientific Computation (CASC)
- Coalition for Networked Information (CNI)
- Education Opportunities (HPC University, SIG HPC Education)
- EDUCAUSE
- EPOC/CI Engineers
- Global Environment for Network Innovations (GENI)
- HPC Systems Professionals
- Midscale Experimental Research Infrastructure Forum (MERIF)
- Minority Serving Institution - Historically Black Colleges & Universities
- Open Science Grid (OSG)
- Quilt (Regional Networks)
- Research Data Access & Preservation Association (RDAP)
- Women in HPC (WHPC)
- XSEDE (Extreme Science and Engineering Discovery Environment)

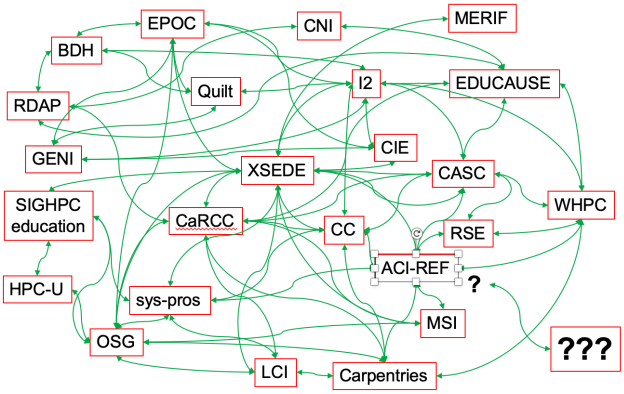

Figure 1: A preliminary, early, and incomplete sketch of the potentially useful interconnections among elements of the RCD ecosystem, with the question marks asking the question: "Who have we unintentionally left out?"

The representatives from each organization prepared a lighting talk to share their organization's mission and to answer the question, "If this ecosystem group could operate as a consortium, what is the number one thing you want from it that you couldn't do on your own?"

Several themes emerged from this question. The members of the ecosystem need to work together to enable communications

and trust ("warm hand-offs"), balance cooperation and competition, advocate on public policy, advance CI preparation, training, and professionalization, establish an authoritative and comprehensive clearinghouse, facilitate access to research computing expertise, and prioritize diversity, equity, and inclusion.

Workshop participants were divided into independent groups and tasked to create visualizations of the RCD ecosystem, discuss what is currently works well, what the challenges are, and what collaborative initiatives might be undertaken to address the challenges that the RCD ecosystem members face. The sketch in Figure 1 was provided as a *straw man* figure to participants during the workshop. At the end of the workshop, two sub-working groups were formed to advance the visualization and clearinghouse activities started during the workshop. The visualization working group was tasked to develop a visualization that captured the current state of the complex ecosystem, and the clearinghouse working group set out to more broadly sample and collect information on organization that make up the RCD ecosystem.

At PEARC19, several of the authors held a plenary panel session to share the developments from the workshop and invite other organizations to participate in our effort to better define the RCD ecosystem.

4 WORKSHOP RESULTS AND DISCUSSION

At the CaRCC Ecosystem workshop, the attendees from a diverse group of organizations involved in the RCD community identified a number of categories of organizations. All had activities or services related to one or more of these categories. Participants also identified a number of characteristic roles that are actors in this community - i.e. researchers or staff with various RCD facing roles. As noted previously, a visualization working group was formed to collect the ideas and draft visualizations created at the St. Louis workshop and to develop those draft ideas into one or two visualizations that would better represent the current state of the RCD ecosystem. As an exercise, the visualization working group examined prior visualizations from the NSF CIF21 document and the 2019 ACI Blueprint. Although created nearly a decade apart, both depicted innovation and discovery at the center of the cyberinfrastructure ecosystem. From here the common themes of connectedness, intersectionality, and boundary-spanning emerged. However, while the entities or "characters" emerged fairly easily, defining the relationships among them proved much more daunting. One popular visualization expressed the ecosystem as a tree, with individual organizations and agencies as the roots, trunk, and branches. Researchers themselves were depicted as leaves being fed and lifted up by the rest of the tree in the same way each supporting RCD organization supports researchers to advance science. Another visualization depicted the ecosystem as a city, with some centralized infrastructure (e.g. water and electricity) and a broad diversity of constituents and roles, each sharing the goal of supporting the lives of individual citizens and families, similar to how the various members and organizations throughout the RCD ecosystem support the process and progress of science.

Also following the April 2019 workshop, a clearinghouse working group formed to create a snapshot organizational listing to document the strengths each group brings to the ecosystem and their connections to other groups to enable stronger collaboration and increased awareness of resources and services across the RCD ecosystem.

The Clearinghouse working group created a survey that was sent to the workshop participants. The questions were:

(1) Organization name
(2) Parent organization (if any)
(3) Web site
(4) Event listings or Calendar
(5) Employment Opportunities
(6) Point of Contact
(7) Brief description of mission
(8) What functions does your organization perform?
(9) What type of organization do you represent?
(10) What are your principal sources of funding?
(11) What are the top 3 strengths of your organization?
(12) List 3 organizations that should be included in the survey.

Initial results and requests for more respondents were presented at PEARC19. The survey can still be accessed at https://tinyurl.com/ci-eco-survey.

There have been 22 responses so far, coming from a fairly even representation of professional associations, non-profit organizations, NSF funded programs, and academic institutions. A summary of responses to the question "What are the top 3 strengths of your organization?" is given in Figure 2. A brief follow-up to update respondents and request identification of links between other members of the clearinghouse led to the addition of 9 more potential entries. The information collected will form the basis for a community resource so organizations can continue to register and update their entries and easily find ways to connect with and leverage the work of other organizations. This platform will also enable those new to the RCD ecosystem to come to a quicker and more effective understanding of the resources and services available.

As illustrated in Figure 1, the RCD ecosystem spans a wide range of organizations representing many different communities interconnected in complex and difficult to understand ways. Moreover, some of the "desired" connections do not exist today. To begin developing an understanding of the strengths and weaknesses of the ecosystem, we posit a simplified version of the ecosystem that groups organizations based on a shared focus or shared goals. Figure 3 illustrates one possible grouping, based on the organizations that participated in the St. Louis workshop. Participants from the workshop included professionals from organizations focused on (1) *high performance advanced computing*, (2) *network connectivity*, (3) *data storage*, (4) *RCD education*, (5) *RCD Advocacy*, (6) *professional development*, (7) *testbeds*, (8) *research libraries*, (9) *best practices/reproducibility*, and (10) *outreach*. Although not represented, it should be noted that some new groups focusing on (11) *resource sharing* among institutions have also begun to emerge. By classifying organizations in this manner it is possible to begin hypothesizing about the existing interconnections between types of organizations.

The lines shown in Figure 3 show our (conservative) hypothesis of where we believe connections exist today, with thicker lines indicating stronger relationships. While one would expect organizations within the same class to be aware of one another and work together reasonably well, it is not clear that sufficient links exists

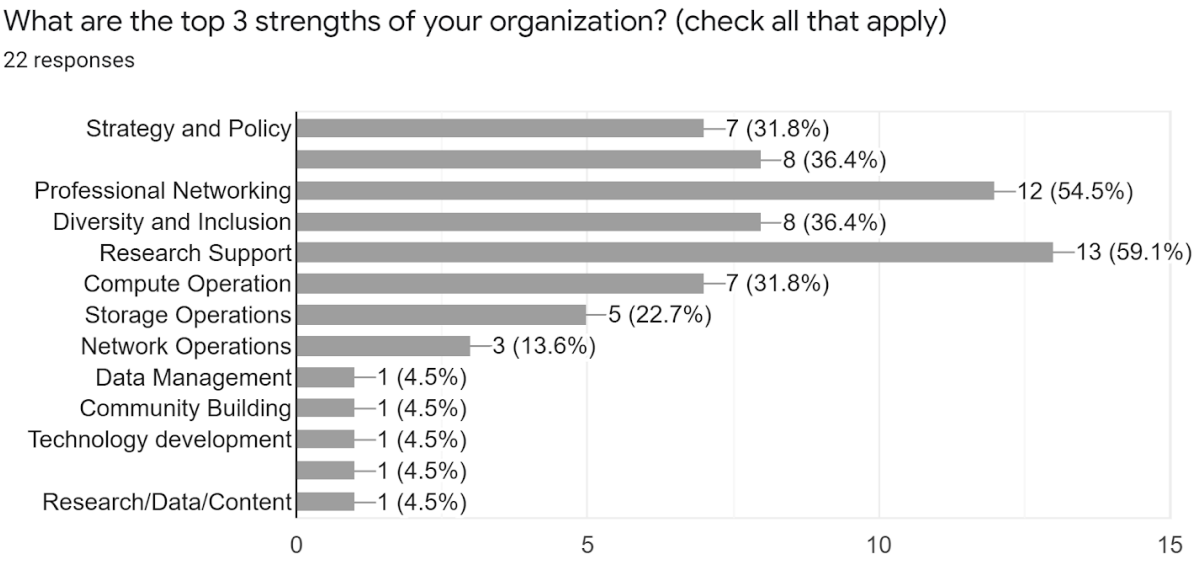

Figure 2: Clearinghouse survey participants identification of top strengths

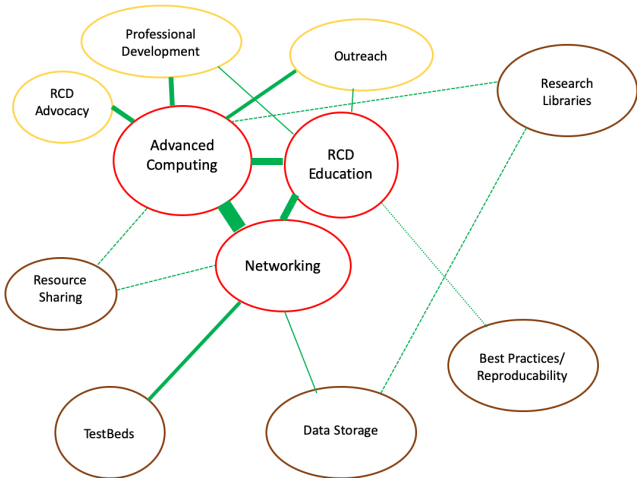

Figure 3: One possible grouping of the organizations that comprise the RCD Ecosystem illustrating areas where connections likely exist as well as areas that may be somewhat isolated and in need of improved connections.

between classes of organizations. Common knowledge – combined with some initial, but sparse, results we have from respondents – seems to indicate that connections between organizations that historically have been associated with HPC in some way are well established, and, in some cases, are quite strong – for example between HPC/advanced computing support organizations, network organizations, and HPC educational organizations. While our estimate of connections is arguably conservative, all indications are that some classes of organizations have relatively few, if any, connections, and appear to be somewhat isolated from the HPC organizations at the center of the figure. This clearly points to an area of potential improvement.

The members of these organizations gather at a host of meetings and venues, some of which are more narrowly focused on a given area of technology, and others of which are designed to be broader and more inclusive. Well-known venues range from the Supercomputing conference series, the PEARC conference series, Internet2 conferences and regional network-centric gatherings (e.g., CENIC, The Quilt, Great Plains Network), to libraries and data-centric conferences such as the Coalition for Networked Information membership meetings and the Research Data Access & Preservation Summit, to vendor-hosted gatherings such as Amazon Web Services' AWS re:Invent and Google Cloud Next. Many organizations of professionals host meetings focused on their membership, including CASC, ARL, RMACC, LCI, and Women in HPC; others are more disperse, tend to meet virtually or locally, and gather face-to-face at other community events (e.g., The Carpentries, Campus Champions).

A not uncommon pattern for conferences and meetings where people meet is to focus on a given area of technology or a segment of the professional community. This is generally not intended to be exclusionary as much as it is a reflection of the human tendency to homophily and a need to have deep discussions among subject matter experts. Even those events designed to gather a broad audience have often had relatively limited participation from some groups in the ecosystem (e.g., PEARC draws relatively few library science professionals). The net impact is that even to the extent

that organizations and domains in the ecosystem genuinely wish to engage more with others, there are real challenges to making this happen.

Given the highly interconnected needs of researchers, there is clear motivation for the organizations supporting RCD to develop stronger ties to broad set of participants of the ecosystem, and to develop a greater understanding of these peer domains. It is clear from the current situation, however, that this will take effort and intention. One of the motivators for this work is to capture and document the current ecosystem and the need expressed by many in the community for a means of discovery of peer organizations (e.g., visualizations and/or directories of some sort).

One important pattern for connecting areas of the ecosystem is to focus on bridging across them, rather than trying to gather them all together. So-called "boundary spanners" are individuals who have roles in several communities, and can act as translators and connectors. Sometimes described as "roving ambassadors…who serve as [a] group's eyes and ears in the wider world," boundary spanners are recognized for the important role they play in situations where people need to bridge or share different kinds of expertise [Cross and Prusak 2002], and to "overcome a boundary and facilitate communication / knowledge flow across it [Long et al. 2013]." Many organizations may (already) have existing or natural boundary spanners, but do not take full advantage of these roles; as part of an intentional approach to building and strengthening connections across the ecosystem, organizations should develop and leverage boundary spanners to partners and peers in the ecosystem. For individual professionals, there is considerable career value in taking on these roles as well: at least one study finds that individuals who take on boundary spanning roles have a significantly increased likelihood of becoming a community leader [Fleming and Waguespack 2007].

5 CONCLUSION AND NEXT STEPS

Many mature and nascent organizations, technologies and professions related to RCD are making significant contributions to the support and advancement of leading-edge science and engineering. The ecosystem concept described in this paper lends itself well to understanding not only what entities are involved in these efforts, but also how the interaction among these entities influences our broader capacity and capability to support current and future research and education. This conceptual framework of interdependent organizations, technologies and professions is also needed to promote better coordination and collaboration among its various members and help to prioritize future developments.

While organizations such as CaRCC, CASC, and XSEDE devote some of their effort to promoting the awareness of RCD-related activities, practices, and resources, a RCD ecosystems-based view must cut across the many organizations that are directly involved in it and is needed to strategically address the rapidly changing requirements of today's science and engineering communities. The RCD ecosystem presented in this paper is admittedly just a snapshot in time of an ecosystem that is in constant flux and will grow to include new members. The relationship among these members will also change to address new research challenges that are difficult to imagine now. An ecosystem-based view of RCD not only gives its stakeholders a way to better navigate this growing and complex network of RCD participants, but allows them to consider more globally how to shape the ecosystem to best serve the needs of its broad set of stakeholders.

Not only is the RCD ecosystem rapidly evolving, but the current characterization of the RCD ecosystem presented in this paper is admittedly imperfect. While great attention was paid to soliciting broad input and participation, we are aware that groups, technologies, and professionals key to some areas of the RCD were missed. From this admission, we hope to use this paper as a general call for participation to help us better characterize the RCD ecosystem. While we freely admit that our characterization fails to provide categories to include all of the RCD participants and lacks a complete description of how these member organizations interact, we are adamant that better characterization of the RCD ecosystem is a critical component of what is needed to accelerate the advancement of the leading-edge of compute and data intensive research. We also argue that characterizing this complex and ever-changing ecosystem simply cannot be done by a single organization; rather, broad participation among the RCD's many stakeholders is required. Furthermore, characterizing the RCD ecosystem should not be treated as a simple "one-off" done every so often. We are seeking a highly dynamic way to capture, describe, and communicate new entrants to the RCD ecosystem that is at least as dynamic as the ecosystem itself.

With the publication of this paper, CaRCC will transition the ecosystems working group responsible for organizing the various events described in this paper to a sustaining group. As a sustaining group, CaRCC members have discussed creating a website that would allow new and existing RCD stakeholders to "self-register" their organization, which could entail describing the broader goals of the organization, and if applicable, how the organization is linked to other organizations in the RCD ecosystem. This information could serve as the data to create real-time views of the RCD ecosystem and could be used to periodically create a snapshot of the RCD ecosystem so that changes to this complex system over time could be better understood. It should be noted that CaRCC is a volunteer organization, which brings with it both challenges and opportunities. The challenges are not unlike those that any other volunteer organization has; namely, how do we sustain our activities when our volunteers' time is pulled in many other directions? The opportunities come from the near limitless set of highly skilled participants who by their participation can represent the entire breadth of RCD stakeholders. The key is that these volunteers must be convinced that their efforts will generate value. So far, this has certainly been the case, which gives us hope that this paper will serve as a springboard for the development of a more complete view and long lasting framework to help characterize the ever-changing RCD ecosystem. Armed with this more complete view of the RCD ecosystem we are better equipped to help to advance the broad goals of new organizations with RCD needs, help researchers quickly find solutions to their pressing research needs, and help university and foundations identify and prioritize investments.

ACKNOWLEDGMENTS

This work was partially supported by NSF grant 1620695 ("RCN: Advancing Research and Education Through a National Network of Campus Research Computing Infrastructures – The CaRC Consortium"). This work was enabled by the contributions of time and expertise from the April 2019 workshop participants and the CaRCC Ecosystem working group members: Amy Neeser, Barr von Oehsen, Bob Freeman, Cliff Lynch, Curt Hillegas, Damian Clarke, Dana Brunson, Douglas Jennewein, Erica Johns, Gail Krovitz, James Deaton, Jim Griffioen, James Wilgenbusch, Jennifer Schopf, Jim Bottum, Joel Cuther-Gershenfeld, John Goodhue, John Towns, Karen Tomko, Karen Wetzel, Kate Cahill, Lauren Michael, Lawrence Landweber, Marisa Brazil, Mary Lee Kennedy, Neil Bright, Patrick Schmitz, Robert McDonald, Sharon Broude Geva, Stephen Harrell, Susan Mehringer, Thomas Cheatham, and Toni Collis; and the Coalition for Networked Information who hosted the workshop.

REFERENCES

Rob Cross and Laurence Prusak. 2002. The People Who Make Organizations Go – or Stop. *Harvard business review* 80 (2002), 104–12, 106.

Lee Fleming and David M. Waguespack. 2007. Brokerage, Boundary Spanning, and Leadership in Open Innovation Communities. *Organization Science* 18, 2 (April 2007), 165–180. https://doi.org/10.1287/orsc.1060.0242

Janet C. Long, Frances C. Cunningham, and Jeffrey Braithwaite. 2013. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. *BMC Health Services Research* 13, 1 (April 2013), 158. https://doi.org/10.1186/1472-6963-13-158