Federating CI Policy in Support of Multi-institutional Research: Lessons from the Ecosystem for Research Networking

Melissa Cragin
mcragin@sdsc.edu
San Diego Supercomputing Center
USA

Ron Hutchins
rrh8z@virginia.edu
University of Virginia
USA

Maureen Dougherty
maureen.dougherty@ernrp.org
Ecosystem for Research Networking
USA

James Barr von Oehsen
jbv9@rutgers.edu
Rutgers University
USA

Michael Zink
zink@ecs.umass.edu
University of Massachusetts, Amherst
USA

D. Balamurugan
bala.desinghu@rutgers.edu
Rutgers University
USA

John Goodhue
jtgoodhue@mghpcc.org
MGHPCC
USA

ABSTRACT

The ERN (Ecosystem for Research Networking) works to address challenges that researchers face when participating in multi-campus team science projects. There are a variety of technical and collaborative coordination problems associated with shared access to research computing and data located across the national cyberinfrastructure ecosystem. One of these problems is the need to develop organizational policy that can work in parallel with policies at at different institutions or facilities. Generally, universities are not set up to support science teams that are distributed across many locations, making policy alignment an even more complex issue. We describe some of the work of the ERN Policy Working Group, and introduce some key issues that surfaced while developing a guiding policy framework.

CCS CONCEPTS

• **Theory of computation → Interactive computation**; • **General and reference** → Validation; • **Social and professional topics → Computing / technology policy**; • **Networks → Network design principles**; • **Security and privacy → Security protocols**.

KEYWORDS

Policy, Collaboration, Research Computing, Federation, Cloud Services, Edge Computing, Core Facilities

PEARC '22, July 10–14, 2022, Boston, MA

ACM ISBN 978-1-4503-9161-0/22/07…$15.00
https://doi.org/10.1145/3491418.3535167

ACM Reference Format:
Melissa Cragin, Ron Hutchins, Maureen Dougherty, James Barr von Oehsen, Michael Zink, D. Balamurugan, and John Goodhue. 2022. Federating CI Policy in Support of Multi-institutional Research: Lessons from the Ecosystem for Research Networking . In *Practice and Experience in Advanced Research Computing (PEARC '22), July 10–14, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3491418.3535167

1 INTRODUCTION - RESEARCH AND LEARNING IN A HETEROGENEOUS DIGITAL LANDSCAPE

Academic institutions are in the midst of fundamental changes to the organization and provision of digital research and learning environments. Systems are increasingly data-enabled, affecting the need for new skill sets and professionalized roles to oversee and deliver administrative, educational and research services. For research in particular, the combination of increasing computational power and high-speed networks provides a foundation for implementation of new methods and access to distant instruments, and supports the coordination of both large and small distributed projects. Beyond the technical challenges, researchers interested in solving deeply specialized disciplinary problems (e.g. finding pieces of a puzzle), or multi and interdisciplinary teams organized to address systems-oriented problems, often need access to cyberinfrastructure (CI) and data resources that are housed across institutions and organizations. Data, software, workflows, and expertise are resources that must be available to users with minimal friction or impediment. Alignment of institutional policy, procurement procedures, and governance practices is arguably the next big challenge for ease-of-access to distributed research resources. This short paper provides an overview of key issues in the context of work undertaken by the Policy Working Group of the Ecosystem for Research Networking (ERN) to gather information to inform the selection of elements for a framework to create federated CI policy.

2 INFRASTRUCTURE-RELATED CHALLENGES / BARRIERS TO COLLABORATION

There is a significant base of literature on the organization of scientific collaboration, team science, and the impacts of cyberinfrastructure on the growth and production of science. Across the social sciences are studies of the structure and function of teams, processes of scientific communication and production, and the impacts of digital and computing resources on the pace of knowledge production. (See, for example, Team Science[1]; Cummings & Kiesler (2005)[2], Stokols, et al. (2008)[5]). However, the role of institutional functions in multi-institution CI federations has received much less analysis, in part, perhaps because much of the work that happens to resolve the challenges is generally invisible. As computing resources and new sorts of instruments (e.g. spectrometers, high-resolution microscopes) have come on-line, at least two kinds of practices are becoming regularized in project design: Distributed, multi-disciplinary team-based activities proliferate, and researchers from a variety of disciplines have access to remote instruments. From a service perspective, the CI community has had to develop policies and procedures for allocation and provisioning, appropriate use, and user responsibilities for data. As local and collective user bases grow and the scale of data increases, policies and terms on data storage and maintenance are revised or added. Often overlooked, though, are policies that apply across systems which will intersect with regulations and procedures at different levels of that system - like academic institutions - and these various sets of policies are rarely designed to address common needs and may not work in concert.

2.1 Multi-institution Arrangements for Shared Cyberinfrastructure

In particular, universities are not set up to properly support team science that spans multiple campuses, often because the right policies are not in place. As projects and production have become more complex, local policies (where they exist) may conflict with an external organization's policies. It is widely recognized that sharing of regulated data involves a complex set of technical, administrative, and legal measures. While the regulations are the same for all institutions, the policies and agreements that support compliance vary from institution to institution. As a result, most new inter-institution collaborations must start with a negotiation of terms. Approval and documentation requirements can be as challenging as the technology, policies, and agreements needed to protect the data, define responsibilities, and limit liability. A standard set of policies and agreements, adopted by all participating institutions, could reduce administrative overhead and allow projects to get started more quickly. It could also make collaborative proposals more competitive by assuring reviewers that data sharing agreements will not become a roadblock. In addition, it could allow smaller and minority-serving institutions to participate on a firmer footing in projects involving regulated data. How do we standardize?

3 ECOSYSTEM FOR RESEARCH NETWORKING

The Ecosystem for Research Networking[3] (ERN, formerly the Eastern Regional Network), was formed in 2017 to address the challenges researchers face when participating in multi-campus team science projects, associated with shared access to research computing and data located within the national cyberinfrastructure ecosystem. The ERN started as a regional effort for two principal reasons: (1) a desire for face to face interactions and physical proximity to and access to shared instruments, and (2) unique characteristics of our region – for example, the Northeast contains eight different state university systems in a geographic area whose size is comparable to that of California, nine different regional network providers, and close to two-thousand colleges and universities of all types and sizes, many of which are under-resourced or under-represented. Though we originally kept our sights to the Northeast, we came to realize that by addressing the challenges unique to the region, expanding our scope beyond the Northeast would not be any more difficult and, in fact, may be less of a challenge in some locations.

Over time the ERN changed its scope when it realized through its interactions with several research communities that the ERN needed to treat physical scientific instruments such as telescopes, research vessels, sequencers, and scanning electron microscopes on an equal footing with research computing, storage, and networking in regional cyberinfrastructure planning. There is a need to access, move, store, and process the massive amounts of data generated by the scientific instruments, and to access the instrument remotely through federated services available to both the researchers and their collaborators. Through interactions with the scientific and cyberinfrastructure communities the ERN learned that these capabilities are often an afterthought when implementing instruments, which can limit the value of these major investments and lead to complications when sharing the instruments and the data. Well defined policies are necessary for maximum use. The ERN is now formulating a federated solution to these campus core services focused on the challenges of collaborative research and education and the supporting policies, enabling the ability to offer their instruments as a service to a broader community, while under local institutional control.

The ERN has organized several working groups to lead the development of necessary aspects of the organization. These include an Architecture and Federation group, the Broadening the Reach group representing under-resourced and under-represented institutions, and a Policy group, which meet independently and in concert to align design decisions and the respective policies. Over the past two years, the Policy Working Group met regularly to put together a policy framework that would facilitate engagement and participation of multiple institutional members with varying cyberinfrastructure resources. A significant part of this effort was to gather expert information on a set of integral components and learn about how they work in practice, particularly those at the intersections with other organizations, technologies, and institutional processes.

4 STRATEGIES FOR POLICY DEVELOPMENT

Policies are a necessary component of collaboration, sharing, and oversight and a key part of a research infrastructure at a university. A sort of framework was needed for discussions to guide decisions on policy scope and specifics. The ERN Policy Working Group focused first on information gathering: We reviewed examples of organizational policies and Charters (ACCORD, Gulf Coast Consortia, Open Storage Network, Open Science Grid, & QUILT), and

engaged with university administrators and technical experts (Rutgers University, University of Virginia, Internet2/InCommon) to identify aspects that must be considered in creating new policies and procedures intended to lower barriers to sharing knowledge, data, infrastructure, and people. The Policy Working Group engaged with the ERN community, and held focused sessions at All Hands meetings to report updates and gather input. In addition, we held joint meetings with the Architecture and Federation Working Group for iterative conversations on progress. Regularizing this process is important to assure that both groups are informed on any changes or emergent dependencies that would impact design, planning, procurement, provisioning, or policy alignment.

5 MINIMUM VIABLE RESOURCE MODEL

As the group worked to frame the relevant facets of action and oversight for the policy framework, one of our meetings was on DataVerse and the FAIR Principles[6], and the kinds of changes happening in research data sharing practices and data services (such as repositories and catalogues). This discussion included the role of technology in policy implementation, barriers and incentives for new practices, and culture change. In 2019, Brian Nosek, Director of the Center for Open Science, published a blog post on Strategy for Culture Change[4] that articulated a framework for organizing action toward collective behavior change. Inside the pyramid (Figure 1) are the layers of responsibility and action, and the accompanying text (on the right) outline the mechanisms for change. Below the pyramid is text listing the layers of the initial policy framework identifying aspects of distributed responsibility or policy action. The mechanisms for change need further development, but include a range of management documents and agreements.

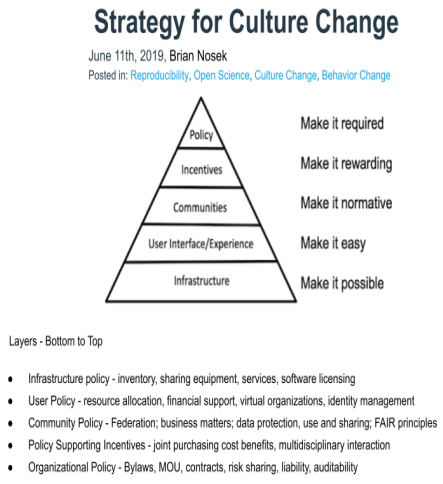

Figure 1: Policy Framework

Albert Einstein famously said that one should, "make things as simple as possible… but no simpler." This quote embodies the concept of the Minimum Viable Resource model. In order for the ERN to accomplish its stated goals of instrument and data sharing across multiple universities, several things have to be put in place:

- A model of instrument that can be shared
- A set of processes for operation of that instrument in a shared model
- A clear method of allocating responsibility and authority among the parties
- A set of policies that allocate risk to these various parties

Development of a minimum viable policy set would include, for example, consideration of, "Who controls the resource and rules for gaining access?" "What is the asset allocation policy?" "What are the criteria for trusted access (e.g. two factorial authentication, ID Proofing)?" And, procedures that will need determination, such as, "What are processes for creating guest accounts?" "What is the a common approach to sharing regulated and unregulated data across institutional boundaries?"

6 MOTIVATION FOR THIS WORK

University leadership is rightly leery of taking on risks, and policy is a means for addressing risk and fostering trust. The benefits must be seen to outweigh the risks when adopting new models of data and resource sharing. One way to show this benefit to risk ratio is by specifying the risks and showing the mitigation strategies. This is a necessary component for hosting data and resources for federal agencies. Compliance systems such as NIST 800-171 clearly outline areas where mitigation is necessary to protect government resources. Following a framework such as this one can help to limit concerns at university leadership level, but these frameworks bring a huge compliance burden with them. Risk management is a necessary part of the technology world today, as is being able to find the minimum amount of compliance-based overhead to protect all the resources we oversee. Even public data we hold, such as weather data should be stored as read-only to ensure integrity and support reproducibility. So, generally we need appropriate frameworks for addressing compliance and plan the work required to protect resources to the level necessary. This is by definition a part of the minimum viable resource model.

7 CONCLUSIONS

Funding agencies are promoting team science; but without policy guidance or proper infrastructure in place, navigating how to access resources and share data across multiple institutions becomes too complex and risks slowing the pace of science rather than easing sharing and access. The ERN's vision is to simplify this process by offering architectural designs that introduce federation, standardization and accessibility by incorporating a comprehensive set of policies and requirements needed for inter-operable, sustainable, secure and compliant systems. These policies and procedures should be clean and clear, and based on the principles of fairness, accountability, transparency and equity. Federated policy should complement local policies whenever possible, by being broad and generally multilateral, and allow the generation of individual bilateral agreements when necessary. The development of a standard terms of use agreement as well as other management documents and agreements could address liability and compliance issues and help to avoid ad hoc negotiation of terms as is often happens today. One solution, for example, is for new agreement templates to be made available to institutions that do not already

have them. The developing policy framework of the ERN Policy Working Group should serve the broader research community by reducing the impediments faced by science teams attempting to access cyberinfrastructure and data resources hosted at remote institutions, organizations and centers.

ACKNOWLEDGMENTS

The authors wish to thank all the members of the Ecosystem for Research Networking Policy Working Group for their contributions to the workshop and the ERN [3] Steering Committee for guidance and support. This work was supported by the National Science Foundation CC* CRIA Eastern Regional Network planning grant (OAC-2018927).

REFERENCES

[1] National Research Council. 2015. Enhancing the Effectiveness of Team Science. Committee on the Science of Team Science. https://doi.org/10.17226/19007

[2] Jonathon N. Cummings and Sara Kiesler. 2005. Collaborative Research Across Disciplinary and Organizational Boundaries. *Social Studies of Science* 35 (2005), 703–722. https://doi.org/10.1177/0306312705055535

[3] Ecosystem for Research Networking. 2021. Ecosystem for Research Networking. Retrieved April 12, 2021 from https://ernrp.org

[4] Brian Nosek. 2019. Strategy for Culture Change. Retrieved April 7, 2022 from https://www.cos.io/blog/strategy-for-culture-change

[5] Daniel Stokols, Shalini Misra, Richard P. Moser, Kara L. Hall, and Brandie K. Taylor. 2008. The Ecology of Team Science: Understanding Contextual Influences on Transdisciplinary Collaboration. *American Journal of Preventive Medicine* 35, 2, Supplement (2008), S96–S115. https://doi.org/10.1016/j.amepre.2008.05.003

[6] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (Mar 2016). https://doi.org/10.1038/sdata.2016.18