

# Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:42:14

PAGE 1

REFERENCE NO: 252

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, [https://www.nsf.gov/publications/pub\\_summ.jsp?ods\\_key=nsf17031](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031). Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

## Author Names & Affiliations

- Charlotte Lee - University of Washington
- David Ribes - University of Washington

## Contact Email Address (for NSF use only)

(Hidden)

## Research Domain, discipline, and sub-discipline

Human Computer Interaction (HCI), Computer Supported Cooperative Work (CSCW), infrastructure studies, cyberinfrastructure

## Title of Submission

Capitalizing formal and informal knowledge of CI development and sustainability

## Abstract (maximum ~200 words).

Abstract: NSF has over a decade of experience developing the sociotechnical support structures for science under the header of cyberinfrastructure (CI). While much has been learned through formal scholarship about CI and through practical know-how, cyberinfrastructure practitioners and scholars alike often find themselves struggling and reinventing the wheel to launch projects, understanding the needs of user communities, and developing infrastructure that is flexible and adaptable to rapidly changing computational-scientific needs. In order to properly support cyberinfrastructure we must also have a deep understanding of the interaction between humans and the various technologies that comprise infrastructure. We call for the development of a center of excellence that will serve as the core of an ambitious research program into the sociotechnical arrangements of cyberinfrastructure, as well as a central resource for cyberinfrastructure practitioners. The center would serve to: 1) recognize the importance of the study of sociotechnical aspects of CI development and maintenance as crucial knowledge that comprises a form of CI in its own right; 2) foster new generations of researchers that investigate sociotechnical arrangements of cyberinfrastructure, and 3) employ staff specialists that coordinate with cyberinfrastructure communities to share organizational models, best practices, lessons learned, sustainability models, and human-centered design practices, etc

**Question 1** Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

A central set of challenges for cyberinfrastructure (CI) are sociotechnical, that is, residing at the intersection of cutting edge computational technologies and the extant organization and practices of contemporary scientific domains. The authors of this submission have been investigating the development of cyberinfrastructure from organizational and sociological perspectives for over a decade, participating in CI

# Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:42:14

PAGE 2

REFERENCE NO: 252

projects across multiple domains. In that time, one outcome is clear: those who are engaged in developing and sustaining computational resources for the sciences -- henceforth cyberinfrastructure practitioners -- find themselves encountering very similar challenges across those projects. Examples follow below.

## Team and Organizational Collaboration

- \*How should teams, workgroups, collaborations, virtual organizations, consortiums, etc be formed
- \*Once formed how can the above types of social organization work together with, through, around, and without CI?
- \*When will ad-hoc collaborations support work more effectively than formal organizations?
- \*How does software developed locally become embedded in larger infrastructural components or within larger infrastructures?
- \*How do development and maintenance practices differ within and between disciplines?
- \*How are different types of CI practitioners involved in development efforts (e.g. designated software developers, data scientists etc.)?

## Forming a Coherent and Inclusive Vision for Practitioner and Domain Community Collaboration

- \*Understanding the domain community sufficiently to build relevant tools and resources
- \*Convincing the domain community to invest in CI development (often perceived as at the expense of domain science)
- \*Getting a CI endeavor off the ground, e.g., the initial weeks and months of stakeholder alignment, visioning, face-to-face meetings, and CS/domain and domain/domain cross-understanding

## Long-term Sustainability of Parts of the Cyberinfrastructure "Ecosystem"

- \*Developing cutting edge computational capacities that will also reward career trajectories in academic CS (etc.) and serve domain communities' scientific needs
- \*Long-term sustainability of: funding, membership, organizational structure, operations and maintenance, etc.
- \*Creating sociotechnical systems that are adaptable, flexible and responsive to ongoing changes in scientific objects, methods, instrumentation, as well as emerging computational capacities
- \*Educating students and CI stakeholders about how to understand sociotechnical systems and about what are common problems, solutions, and interventions when a solution is not readily available.

Throughout our research of CI we have seen these, and other kinds of tensions crop-up in virtually every endeavor. These challenges are thorny, but are not insurmountable.

However, we have also seen that practitioners in each CI endeavor find themselves, first, having to discover these challenges for themselves, and thereafter inventing in-house solutions. While every CI endeavor will encounter unique challenges, we assert that there are structural and general patterns to these difficulties, and that they can best be addressed by supporting a scholarly community dedicated to its systematic inquiry and to the sharing of formal and informal knowledge across CI projects.

**Question 2** Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

We call for the creation of a center of excellence that will serve as the core of an intensive research program into the sociotechnical aspects of CI design and sustainability, and that will serve as a clearinghouse for organizational models, best practices, lessons learned, sustainability models, and human-centered design practices.

It is crucial that this center be both a research endeavor and service center for CI practitioners. Investigating the sociotechnical activities of CI is itself a scientific endeavor, an ongoing research program with funding awards, publications, and academic researchers at all career stages. In our experience, the best way of developing knowledge of sociotechnical aspects of CI is to do so by participating in those projects themselves through research that engages the community in a variety of ways. For over a decade, we have been developing

# Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:42:14

PAGE 3

REFERENCE NO: 252

methods to conduct multi-site investigations with the aid of students and postdocs. Most of these projects have been undertaken on a small, short-term, and local scales. In order to discover more impactful, generalizable and transformative knowledge, a center of excellence is required that can support a research scientists and staff associates to ensure continuity of institutional and methodological knowledge to enable projects of larger scale that are necessary to answer hard questions. This center of excellence will also provide a reliable set of staff to act in consultation with CI stakeholders. Together researcher and staff would provision knowledge and information to those who are trying to improve the effectiveness of the larger social and organizational endeavors that are driven by CIs.

Only a fraction of the projects that can be called cyberinfrastructure are funded by the office of cyberinfrastructure. Many, perhaps most, of CI is supported by the domain directorates of NSF and other state agencies of science funding, as well as consortia, professional associations, private foundations or universities. However, OCI is in the position to support CI generally, across a variety of funding bodies, by creating of a center that will catalyze research and act as a clearinghouse for collating and sharing formal and informal expertise. To date the sociotechnical study of CI has either been funded as a small add-on to an infrastructure development project or it has been funded as its own project as a short-term endeavor. However, infrastructure is a long-term and longitudinal endeavor. If we want to solve the big problems that cross-cut the sciences and that address the reality of CI as also cross-cutting multiple layers of technical infrastructure as well as multiple layers of social organization, then we must fund this ambitious endeavor accordingly. Substantial advances can be made through aggressive pursuit of a more ambitious program. We need larger scale "human infrastructure" to study and support cyberinfrastructure as a uniquely complex, large scale sociotechnical endeavor.

**Question 3** Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

While the most important mission of this center is conducting the cross-cutting investigations of sociotechnical systems that could be truly transformative, there are also impactful short-term gains that can be realized with the creation of a center that includes staff members who are specially designated to engage community members.

Staff could collect and assemble best practices across CI projects. "Low hanging fruits" are available that would provide a useful set of ready-to-hand lessons-learned and best practices for CI stakeholders. This is not the revolutionary work that most PIs studying sociotechnical systems are pursuing, but is important work that would have a substantial impact in a short time frame. Staff could also engage community members and run trainings and workshops for CI stakeholders. Once best practices have been assembled, staff could develop metrics for evaluation and assessment for before and after the adoption of best practices and then could discuss with stakeholders the findings and next steps.

Workforce development activities would fill a large existing gap right now. Domain scientists, data scientists, software developers, and others engaged in CI-intensive knowledge creation are not trained to recognize collaborative structures and lack the tools to think sociologically about how their own practices and artifacts affect those of others. The ability to understand and anticipate the ramifications for working with collaborators or within a larger organization when they change a practice (e.g., software is no longer appropriate) or an artifact (e.g., software does not meet all the stakeholders needs) will be crucial for more effective, efficient, and sustainable CI. Currently CI practitioners have to learn all these lessons the "hard way" or they never learn them at all, requiring other people to clean up the mess downstream. This "mess" often takes the form of expensive redesign and redevelopment of software tools and systems, poor management of human and technical resources, or failure of adoption or use of new CI or components of CI. We need to foster the training of people who cultivate and have the knowledge and skills to map, analyze, and guide others through, a complex landscape or organizations and the people and things that make them effective. A center of excellence will coalesce and invigorate these efforts.

## Consent Statement

- "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

# Submission in Response to NSF CI 2030 Request for Information

**DATE AND TIME:** 2017-04-05 14:42:14

**REFERENCE NO:** 252

**PAGE 4**

---

---