Negotiating Modalities of Cyberinfrastructure Use in NSF’s XSEDE project

This project focuses on tensions within the strategy and utilization of the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE), and how this reflects the shifting makeup of the science community in the United States. XSEDE is a project funded by the NSF to manage the access and service delivery of computational resources in support of basic science. XSEDE provides a unified accounts and accounting model, software packages, training and documentation for users of computational facilities at a number of *Service Provider* resources (supercomputers, high performance storage systems, and visualization systems). Because of this arrangement, the framework for access and service delivery can remain, while Service Providers may change as funding for systems is awarded and expires. This segmentation of service delivery from management functions allows for a flexible relationship between users and the overall cyberinfrastructure resource, but also means that the institutions participating in XSEDE are collaborating to build a service delivery organization while competing for other NSF awards. XSEDE as a large-scale cyberinfrastructure provides computational resources in support of basic science, allocated by a committee of peer scientists that determine viability of analyses in the cyberinfrastructure environment. As the NSF continues to emphasize its agenda to diversify the range of institutions it supports and to broaden the impacts of the resulting science, the mission it sets forth for the XSEDE project to pursue is to continue to provide high-level computational services to established expert communities which are traditional users of these types of resources. At the same time, the project is required to provide resources for under-served institutions and disciplines which previously had little computational leaning. Within the XSEDE project, this represents an opening of the infrastructure from the communities that have a long history of being highly aligned with computational approaches -- high-energy physics, materials science, chemistry, and climate science -- towards life sciences, social sciences, and humanities.

I contend that this transition within the XSEDE project reflects a larger transition in the scientific community. Recent developments in science policy focus on the recent advances in computational methods for solving problems, establishing a “third leg” of science in computational analyses (Reed, et al, 2005), in addition to theory and experimentation (swiftly followed by the "fourth leg" of data-intensive analysis). Disciplines which heretofore have had limited computational capabilities are beginning to make use of simulation and large data sets in order to conduct analyses, in part spurred on by the promise of additional resources for making use of computational techniques. Not all observers frame the rise of computational capabilities to new pillars of science, noting that the decomposition of problems into tractable measures of work for human computation has been a part of scientific analyses since ancient times (Vardi, 2010). This shift in methods of inquiry has implications for the types of analyses that can be conducted, and the conclusions that can be reached.

XSEDE is under two sets of competing tensions: firstly, to provide next-generation computational services to highly computational users at the limits of scale and secondly, to provide a grid-like computational service to researchers who do not have sufficient resources at their home institution. It seems clearly possible to satisfy both of these demands in a technical sense, but the realities of serving two types of user bases quickly become complex. In this way, XSEDE is a theater in which the tension between the NSF's two objectives is negotiated, and users, staff, and management determine the modalities of use of computational infrastructure.

As the XSEDE project strives to provide a ``general-purpose cyberinfrastructure'' for support of basic research, I believe it is necessary to ask `general' for whose purposes? As the tension between traditional and newer ways of using cyberinfrastructure is negotiated, who benefits and who loses?

How does the availability of resources and the emphasis on computational methods shape the inquiries that are made and the conclusions they bring us?

This research project attempts to answer the question: “What are the processes by which XSEDE as an organization attempts to provide resources for both of these communities, what are the challenges that new modalities of computational analysis face, and what implications does this process have for the broader research community?”

I plan to address this question by conducting both quantitative and qualitative analyses. The quantitative work will focus on the use of XSEDE accounts and accounting data, as well as publications data, and geolocation data to understand the usage of XSEDE resources by different fields of science, institutions, and the demographic breakdown of the XSEDE user base. Publication data will be analyzed in order to understand coauthorship relationships and to link publication to usage of XSEDE resources. Qualitative work will be based on participant observation of the XSEDE project activities, including management meetings, conferences, training workshops, as well as interviews with XSEDE staff members and leadership, users, and professionals from other cyberinfrastructure projects such as the Open Science Grid.

Quantitative work will provide a basis for describing the makeup and use of XSEDE, including identifying groups of users by fields of science, location (EPSCoR, Minority Serving Institutions, and R1 universities), and by demographics, and linking users together by coauthorship networks. The interview and observational work will provide thick description of this complex project and the interests, motivations, and activities of multiple players within the project. Both types of analyses will inform our understanding of collaborative resources for the support of research and the flourishing of Big Science projects which rely on many collaborating researchers working together in order to solve difficult problems.

I expect to find that there are many different groups in XSEDE with different values, interests, and usage, and I am hoping to see what level of allocation and usage these different groups are able to make, in order to make a distinction between the resources available to Big Science projects versus that of other groups, along with some of the research productivity of these multiple groups. Understanding the relationship between resources and productivity for these different types of projects is essential to the definition of effective science policy initiatives which direct the basic research activities of the nation.