

Executive Summary of XSEDE Advisory Board Meeting, April 16-17, 2019

Meeting Date: April 16-17, 2019

Meeting Place: Embassy Suites Hotel, Rosemont, IL

***Preface:** This was an extended face-to-face meeting. Included was a discussion of possibilities for a follow-on award and presentations from each L2 area on their PY9 plan highlights. Agenda, meeting notes, and slides are available for XAB members at*

<https://confluence.xsede.org/display/XT/XAB+2019+April+Face-to-Face+Meeting>

Summary of meeting comments and XAB suggestions

Members voted to unanimously approve the February 2019 XAB meeting summary.

Post XSEDE 2.0 Possibilities

John was invited to NSF to share his perspectives about possibilities for what should happen following XSEDE 2.0 and had a productive conversation, but following the government shut-down he is not sure whether the conversation has continued or not.

In February, Manish Parashar (OAC Director) circulated an email requesting feedback to a draft blueprint for how OAC might move forward with Cyberinfrastructure. John will share this document with the XAB so they can provide feedback if they wish. There is concern that NSF will not have a new award in place before the end of XSEDE 2. It would be appreciated if XAB members would encourage NSF to move this forward. While XSEDE applauds Manish for reaching out to the community, the process seems stalled.

The board noted that XSEDE has created a virtual community and social aspects of Cyberinfrastructure that are essential to bring the community together, which is something NSF hoped to accomplish. XSEDE needs to better articulate this so it can be acknowledged and appreciated by NSF. If certain components of XSEDE go away, that would make the whole project fragile. XSEDE spent the first part of the project thinking about strategic goals and accountability and has that part down, but other accomplishments are equally significant and should be shared. This argument could be framed through shared values. NSF is very community driven and depends on community members to serve as review panelists. These people read project reports and make decisions. XSEDE needs to consider whether panelists can answer questions about the overall value of XSEDE from reading the reports and whether there is enough evidence that this is needed in the community. The reports currently do not provide an easy way of drilling down or convey clearly enough how much this overall cohesion makes a big difference. The project should consider how to accomplish this without making the reports longer. John noted that the value of the whole vs. the sum of the parts is something the project struggles with adequately expressing, and he is considering rewriting the first section to try to articulate this. The board stressed the importance of strategic plans, accountability, and shared values of the community. The current report structure does not indicate the project's tactical approach toward achieving strategic goals even though there was much time spent on that. This should be addressed in the report in a paragraph or two.

John noted that using KPIs to make decisions is working, but he agreed that XSEDE needs to talk more about the process used. The board noted that using the idea of shared values to talk about the process would be beneficial, and to focus on people and how XSEDE serves the community as key values. The board also noted that use of KPIs are effective for internal use, but those outside the project don't understand what they mean or what hitting the targets show the project actually achieved. XSEDE needs to assess real impact for researchers, show that researchers who have

XSEDE support have greater impact than those who do not, and show how support for campuses really impacts those campuses. John noted that XSEDE currently only has anecdotes to tell this story. It is important that reviewers hear the story about where there has been community impact and the delta between TeraGrid to XSEDE1 and XSEDE 1 to XSEDE 2. The project should show the pieces that were not in place but now are, provide as much context as possible for reviewers since some will come in without perspective or history, and focus on these stories instead of KPIs.

It would also be valuable to show XSEDE's value to providing campuses with compute cycles. John noted that the nature of the reports is that they are snapshots in time. The project is working with IRIS at the University of Michigan to determine the economic impact of XSEDE, which is important information for NSF to share with Congress. The project is also planning to do several longitudinal studies to see the project impacts and how things evolve over time to help us better tell its story. XSEDE could add updates on these efforts to the annual reports to show trends over time to better tell its story. There are also things XSEDE talks about during the review that don't show up in the annual reports, such as amount of dollars of federally funded research that are supported by XSEDE, which is in the billions of dollars. The board noted that many campuses do not have research computing support so XSEDE provides that for them. This should be highlighted. The overall value of science accelerated is greater than the investment. NSF wants to be convinced that their investment in XSEDE is better than any alternative.

It was noted that many people use XSEDE training who otherwise wouldn't have access to cluster computing. This indirect impact may be bigger than the direct science impact. John noted that he includes in his opening presentation of the review the number of people who use the XSEDE User Portal, and that approximately 25% of them do not have allocations so are using it for other purposes.

The board noted that it is incumbent on XSEDE to demonstrate to NSF that it is less complex (i.e. more efficient and streamlined) to be done by XSEDE than anyone else. Reading the annual report is complicated, and having an enhanced version of an org chart to show interconnectedness of the WBS areas could be helpful. This is presented to the review panel but could also be included in the report. Staff noted that there are mixtures of people from a variety of institutions that perform different functions. XSEDE needs to show that it's not modular and better show interconnectedness. The board noted that the current Track 2 proposal is worth \$22M, so the project needs to convey that XSEDE is worth more than one of those resources. John mentioned that the project is developing ROI and cost avoidance arguments, which is a new approach. XSEDE is working to bring together the ROI work and the work with IRIS to independently determine the same kinds of things. This still misses the human side though. For example, ECSS was originally part of the hardware award, and when that changed, the project characteristics changed. The story of why it is better to do it this way than what had been in place before is important and should be in the report. XSEDE has various people from various institutions helping someone use a resource from yet another institution. It is important to demonstrate what costs are avoided by doing this together. NSF needs to clearly see the success of the project. Producing ROI information takes time.

The board noted that if NSF were to go to separate awards they would likely still need to have a coordination award, which would be a variation on the TeraGrid model, which did not work, due to some issues John described to the board. John noted that the Service Providers (SPs) are integral partners to XSEDE, they are not funded through XSEDE. They are required by NSF to provide services to XSEDE. TeraGrid had a smaller coordination award (\$9M over 5 years), and more funding flowed through the SPs. There were unfortunate side effects to that model, so corrections were made. John headed the TeraGrid resource provider forum, and all PIs of hardware had veto

power in decision making which made it nearly impossible for any decisions to be made because everyone had to agree, which rarely happened. When the model was changed, that changed everyone's behavior, and issues with SPs have always been able to be resolved without escalating anything to NSF. XSEDE has built trust within the community which makes it easier to move things forward. Subaward partners are chosen by XSEDE and funded under a single award. We have gotten to a far better place with this model. When a particular skillset is needed on the team, the project looks to its partners to make new hires of the necessary talent. XSEDE has a coordination role in the overall ecosystem, including OSG, Software institutes, etc. This needs to be expressed in the report. There are also many other funded projects that collaborate with XSEDE that XSEDE doesn't receive funding for. This shows how XSEDE is a nexus in the community for many other activities.

PY9 Plan Highlights

The meeting then shifted to each L2 area sharing highlights of their PY9 plans. Slide decks can be found through the link provided in the preface.

CEE PY9 Plan Highlights

Regarding KPIs, the board noted that it would be more informative to show the trend of how the project has done over time than providing the single data points. Staff responded that we do show such trends in slides presented to NSF at the review. It was also noted that the report includes text in the introduction that defines KPIs and how the project uses them. There is a lot hidden behind them by design, but overall they are indicators of how the project is doing. If there is improvement in an indicator, the project is moving the right direction. We have removed much of this information from the report in an effort to shorten the doc. The spotlight chart provides a quick view of how the project is doing.

The board would like to see the breadth of training offered discussed, as many campuses augment available training with XSEDE training. They also feel it would be important to show how many individuals participate in XSEDE training, whether they are moving upward in their knowledge, and as they gain advanced knowledge they move more advanced code integration. Staff responded that we can report on the products/outputs we see from them if it's done within XSEDE User Portal (XUP). Kelly noted that she will add information regarding the breadth and depth of training over the years and how the project has evolved in this area.

It was noted that workforce development compliments more formal training offered at universities and includes many skills that are not taught in the classroom. The project is seeing an increasing number of college students participating, which allows the project to have a deeper engagement with universities and provides students with access to training without addition tuition costs. Future training topics are gathered from a variety of sources including the User Survey, Service Provider user communities, and technologies coming. Universities of all sizes take advantage of the training opportunities. XSEDE is at the forefront of science and tools being used, and efforts in Workforce Development allow community members to be trained with the newest technologies as they emerge. The board noted that while offering training for students is important, the project should also focus on offering similar training to those already in the profession so they are equally educated about the latest technologies and don't get stuck in doing things the same way they always have. The project should also reach out to people trying to re-enter the workforce after time away. They board suggested that the project offer a summary of training that people can share with their supervisor.

Broadening Participation's work with student programs has received a best paper award, and the team is often asked for advice by other organizations who want to build similar programs from

ground up that encourage students to go back into their communities and address community needs. That is, teach data science in terms of local, national or global community issues. Kelly noted that without XSEDE, she wouldn't have had access to such a diverse group of students or had the conversations to lead to creating the Advanced computing for social change program. Relationships developed through XSEDE started the cascade for this program, and that led to creating Computing4Change, and that inspired the NSF INCLUDES program. Involvement with XSEDE provided access to the right collection of people and the funding to allow us to try things.

John noted that the project has posted a page on the website with information about how other projects can collaborate with various aspects of XSEDE that can be pulled from when applying for grants: <https://www.xse.de.org/about/collaborating-with-xse.de>. The community has responded very favorable to having this information readily available.

RAS PY9 Plan Highlights

The main focus of the conversation about RAS's PY9 plan was the percentage of allocation proposals that are denied and how the project might work to improve this percentage. Currently there are approximately 70-75% successful allocations. Many of the projects that are declined just failed to include an adequate code performance & scaling document (a requirement), and staff is working to help improve this by providing templates to guide people through the process. There are three reasons that proposals are typically rejected: 1) there aren't enough available resources; 2) the proposed experiment design is flawed (typically results in reduction); and 3) technicalities such as failing to mention resources available and having a poor code performance & scaling document (typically results in rejection). If a proposal is just above the threshold of being good enough to pass at low level, the tendency is to reject and ask them to correct & resubmit. This is concerning because many people who are rejected never return. The board questioned whether bringing in new reviewers has had any influence on the process. Dave has pulled data on every reviewer, including how many meetings they attended and how many proposals they rejected. Newer reviewers are less likely to recommend rejection, but they are often persuaded by the panel to reject. In particular, there is small core of newer reviewers who are associated with a 50% rejection rate, which is concerning. It was noted that while there isn't a specific training for new reviewers, there is a guide provided to them. Formal training could be helpful. RAS has asked the community whether receiving some (reduced) allocation is better than having none, and the response was clear that they would rather have reduced allocations. The board suggested that RAS consider giving allocated projects priority but allow others to submit jobs with a lower priority. Dave noted that this would have to be coordinated with the SPs, and they are fully subscribed. This would allow people without allocations to use resources as long as they're not blocking those who have been allocated.

The board questioned whether any particular demographic group is more often rejected. Dave responded that while he hasn't noticed anything out of the ordinary for any particular group, we only have this information on about 10% of the proposers (many choose to not provide demographic information), and otherwise have to depend on MSI & EPSCOR data. The board recommended that this information be collected separate from the submission process in an effort to gather more demographic data.

It was noted that users are required to agree to an acceptable use policy to use any XSEDE resources. There are security measures in place (i.e., someone from North Korea can't get onto any XSEDE systems). 20% of users are from outside the U.S., but usage from this group is low. All users must be associated with a U.S. institution.

ORCID IDs will soon be integrated with XRAS so allocation information can be added to user profiles. This feature will go live in the June timeframe.

Other highlights mentioned were that XRAS has added a dashboard of activities that need to be processed, which makes it easier to work with submissions. The XSEDE accounting system predates TeraGrid and doesn't function at the scale needed by the project. The team is working to improve this with a goal of having it ready for PY10 and whatever follow-on award there will be.

Ops PY9 Plan Highlights

It was noted that GridFTP is related to Globus online. The project pays a lot for this, so it is looking for alternatives. GridFTP doesn't take advantage of newer technology. The board noted that Open Science Grid is facing the same issues, so there may be an opportunity to collaborate. Ops works to track new developments as new technologies come along. The board noted that this is an area where engaging with the user community is important. Networking & security do benchmarking, but this doesn't address usability. Greg noted that he will give this feedback to those teams and encourage them to think about better ways to engage users. Greg noted that if the XSEDE award was split apart, managing Operations would be incredibly challenging. It was noted that other areas of the project would have trouble as well if Operations wasn't centralized.

XCI PY9 Plan Highlights

CRI has exceeded expectations with the number of schools they have gotten up & running. This ties back to RACD because the next thing these campuses will want is containerized software. The board questioned whether it would be possible for XSEDE to play a role in taking all the components of the Cyberinfrastructure for Sustained Scientific Innovation (CSSI) program and build a coherent national or international CI ecosystem. Dave responded that this is possible but would expand beyond XCI. XCI has become the go-to place for software and solutions for the ecosystem, and XSEDE is the glue for the ecosystem. The board suggested that XCI create a map or visuals to show all campuses that have used and depend on these services. Dave responded that CRI and RACD could both do this. It was noted that XCI is a resource, and XSEDE as a whole provides the coordination. XCI experienced significant budget cuts from XSEDE1 to XSEDE2. Dave noted that these cuts forced the team us to be more focused and efficient, but also made them realize that they're not effectively telling the story about everything they're doing. In shifting to XSEDE 2, the project moved to a lighter weight process. By year five of XSEDE1, there was a foundational infrastructure that we could build on with a lighter weight process and still accomplish a lot on a smaller budget. Depending on what the follow-on solicitation looks like, XCI could potentially be expanded in an effort to build the community ecosystem. Currently, XCI doesn't have enough staff to fully support helping campuses get started. Interest in Aristotle¹ is growing fast, and if we ignore this we will lose an opportunity. XCI would also needs an increase in staff to focus on more use of cloud and providing containerized versions of these codes to give flexibility. If we don't increase staff in these areas we'll fail to fulfill future needs, and the community will lose. The board noted that the project should look for more partnerships in the community instead of expecting XCI to do everything, and while community support is important, it is not as effective as having someone take responsibility. XSEDE has accepted and delivers on that responsibility. We do well because we're funded to do so, and we can't expect to fall back on an unfunded group to do this work. For the first time we're bringing forth software that lets us bridge bi-directionally. We evolved to this for a reason. To tear it apart would be tragic, and users would suffer.

ECSS PY9 Plan Highlights

The board questioned why XSEDE is putting significant funds into Gateways² when there is a Gateways Institute. Staff responded that there is a great deal of effort for this that requires a division of labor. Gateways lived in XSEDE for a long time until the Institute was formed, and when

¹ <https://federatedcloud.org/>

² <https://www.xsede.org/ecosystem/science-gateways>

that happened there was an assumption that XSEDE would play a role. We have been careful not to duplicate efforts. There is a benefit of cross-pollination of Gateways folks working with other aspects of ECSS. Especially as the software ecosystem continues to evolve. There is a benefit to sharing people and expertise.

The board noted a concern with the lack of gender diversity among ECSS leadership. It was suggested that ECSS be proactive and highlight efforts they're making to change the diversity.

The board questioned whether the team has considered offering extended support for cloud computing. Staff responded that they submitted a project improvement request to get more resources for such support. A lot of effort is put forth to get in front of new developments.

The board asked how ECSS works with CEE on expansion of its user base of underrepresented communities. This is done through Extended support for education, outreach, training, as well as a parallel effort by NIP to reach new communities. The project might consider demonstrating how people in these various areas represent different institutions to show interconnectedness. Staff responded that they will consider creating a heat map representation to help illustrate this visually.

The board noted that the amount of money spent on ECSS to make researchers more productive is small compared to money they received for their research. There is a question in PI exit interviews that assesses the value they gained from XSEDE ECSS. Many say it wouldn't have happened without ECSS. Most say working with ECSS saved them over a year of time they otherwise would have spent. When ECSS undertakes projects, they meet with the PI and put together a work plan to be sure the services are being used effectively. Effort is made to ensure that ECSS staff can have maximum impact.

The board noted that NSF won't provide long-term support to sustain software. XSEDE tries to put things in place so the community can sustain it.

The board noted that it is confusing to move KPI targets from /qtr to /year. Phil responded that in some cases we have it better to look at certain KPIs over a full year instead of dealing with the fluctuations you sometimes see from quarter to quarter.

The board noted that while growth of 20% may not seem like much, this significantly enables work for people in these non-traditional communities. There is a larger investment than just cycles.

ECSS is putting more emphasis for staff on developing skills in order to stay out of the forefront.

The board noted that XSEDE is unique with this deep integration of ECSS with researchers. Some research projects hire expertise for the life of their project, but XSEDE ECSS staff learn multiple tasks, which makes them stronger.

The board noted that if the project is successful in recruiting new projects, it will run into manpower issues. Scalability might be increased by documenting best practices and doing a video series. Phil responded that ECSS does some of this with its monthly symposium series, but we haven't taken the next step to distill into a format more accessible to the community. This would be a good legacy product.

The board suggested that the project consider a reward mechanism so staff are encouraged to submit to the Technical Report Series. We might select a "Best Tech Report of the Year", for example.

The board suggested that instead of assuming that a 5% budget reduction equates to 5% cut in projects, ECSS might take a retrospective look at what it wouldn't have been able to do last year if they had lost 5% of funding.

It was noted that ECSS benefits from leveraging across centers and having a bigger pool of people and skills to draw on. ECSS is heavily integrated with other areas of the projects including XCI

(supporting new communities, software, cloud), RAS (reviewing), and CEE (provides trainers). There would be a significant increase in latency and overhead if the project were split apart. The board noted that without XSEDE ECSS, the community would lose an important service. It was suggested that the project create a common slide to show how each of the L2 areas are dependent on one another to demonstrate that XSEDE can't be easily broken apart.

Program Office PY9 Plan Highlights

The board noted that when communicating about XSEDE, it is important to include the word "services."

Options being considered for longitudinal studies were shared. Board members felt that options 1, 4, 5, 7, and 8 were most important. [Please see the presented materials for descriptions of these.]

The board suggested that the wording of the KPI related to PIFs needs to be revisited.

The board questioned whether there is concern about the performance of the KPI on wiki satisfaction. Lizanne responded that scores for this started in the 2s [out of 5], but following improvements to the wiki and adoption of other communication tools, scores have improved and continue to climb.

The board suggested that the project shouldn't set targets much above a 4, and that the project should avoid adding .1 each year.

Lizanne noted that compared other work she has done, XSEDE has very good ratings overall. The board suggested that those reading the report would find it more interesting to see how many KPIs score above a certain target than looking at each KPI individually. Lizanne noted that we do this already.

Wrap-up

The board noted that this generation of students are digital natives who easily find resources online. Given that they are savvy about teaching themselves, XSEDE may need to make adjustments in order to stay relevant. Usage from new communities who need more hand-holding is increasing at a rapid rate, which increases the demand for XSEDE. XSEDE needs to be there for the community or someone else will be.

The board noted that the newly awarded Track 2 machines will use XSEDE but also go 2+ years beyond the end of XSEDE2. Some of these new machines could place different demands on XSEDE

The board discussed possibly writing a statement about their beliefs of the needs following XSEDE2 in order to meet the needs of the community. CaRCC & CASC could echo this statement, and CNSF could also be asked to write a letter.