Report of the NSF CyberBridges Workshop:

Developing the Next Generation of Cyberinfrastructure Faculty for Computational and Data-enabled Science and Engineering

June 25-26, 2012
Arlington, Virginia

Thomas Hacker
Purdue University

Suzanne Shontz
The Pennsylvania State University

Table of Contents

Executive Summary

The first workshop for the NSF Office of Cyberinfrastructure CAREER awardees was held June 25-26, 2012, in Arlington, Virginia. Twenty-five attendees and five keynote speakers attended the workshop. The attendees, who were funded by the NSF BIO, CISE, ENG, HER, MPS, and OCI directorates, were selected from 23 institutions. Five keynote presentations were given by nationally recognized leaders in fields relevant to the use and development of cyberinfrastructure in science and engineering research. Each keynote presentation was followed by a discussion session with workshop attendees.

The workshop provided a venue for CAREER awardees to interact and to develop new collaborations with leading researchers and other CAREER awardees. As a result of this workshop, 55 potential new collaborations were identified by attendees.

The workshop provided many opportunities for discussions among attendees and speakers. We received many comments from the attendees. A distillation of these comments can be summarized in several observations.

Observation 1. The attendees felt that there is a lack in the number of interdisciplinary solicitations that bridge the domain sciences, computer science, and cyberinfrastructure.

Observation 2. The attendees suggested that the CAREER proposal review criteria could be expanded to encourage the development and dissemination of open-source software and data as part of the proposed project.

Observation 3. The attendees described an unmet need for a central cyberinfrastructure that could be used by CAREER awardees to disseminate open-source software and data (an example of this type of system is the NEEShub).

Observation 4. To assist early-career researchers in developing their research programs and where and how to seek federal sponsorship for their research, many attendees described a need for clear guidelines to CAREER awardees in the science and engineering areas covered by NSF programs and solicitations, and when researchers should approach other agencies, such as NIH, DoE, etc.

Observation 5. The NSF CyberBridges Workshop provided a productive venue for CAREER awardees to meet, network, and explore integrative and joint research and education activities; and to allow senior researchers to provide career advice to young researchers.

Observation 6. There is a need to create more core solicitation programs with a well-defined submission schedule to allow researchers to better plan in advance to submit proposals to a specific program, rather than the need to respond more quickly to singular solicitations that are not part of a recurring program.

Workshop Overview

In 2010, the NSF Office of Cyberinfrastructure launched a new Faculty Early Career Development (CAREER) program to support investigators working on interdisciplinary research in cyberinfrastructure and the application of cyberinfrastructure to science and engineering research. At the time of the workshop, approximately 50 CAREER grants had been awarded to researchers. To bring together the community of OCI CAREER awardees, a workshop was held June 25-26, 2012, in Arlington, Virginia with several goals: (1) encourage networking and discussion among awardees; (2) provide a forum to facilitate the discovery of new synergies and connections among researchers from the community; and (3) provide inspiration and motivation for new research through a series of keynote presentations by leaders in the areas of Grand Challenges in Cyberinfrastructure, Big Data, Computational Science, High Performance Computing, and Visualization. The workshop provided networking opportunities to seek out and gain potential collaborators, and included a poster session that allowed poster presenters to solicit additional interest from attendees.

Outcomes of the NSF CyberBridges Workshop

There were several common conceptual themes that emerged from keynote speaker presentations and the ensuing discussions with attendees that were led by keynote speakers. As workshop co-chairs, we identified ten key subthemes based upon these themes. Five of these topical areas represent thoughts on ***interdisciplinary opportunities*** in the computational- and data-enabled science and engineering and cyberinfrastructure fields, whereas the other five **address *challenges in these fields***.

A. Interdisciplinary Opportunities:

1. **Interdisciplinary initiatives and universities:** Although no university structure is perfect, both graduate training and the tenure and promotion process can be more effective and efficiently structured when faculty activities can be supported and integrated within an existing interdisciplinary center or institute.

2. **Interdisciplinary initiatives at NSF:** More interdisciplinary proposal initiatives are needed. Some existing multidisciplinary or interdisciplinary proposal ideas fall through the cracks. Also, the interdisciplinary/multidisciplinary solicitations are always changing, so it can be difficult to be able to count on solicitations being in place. An improved process for soliciting and gathering feedback from program directors across NSF directorates is needed for interdisciplinary solicitations. Clearer and more coherent information on interdisciplinary initiatives would help proposal teams to more effectively process and respond to program director comments.

3. **Academic-Industry Partnership:** We discussed the need to develop better linkages between academia and industry to encourage technology transfer to industry and to promote the development of sustainable cyberinfrastructure components.

4. **Collaborations:** Dr. Chris Johnson offered his model for developing successful collaborations. Prior to initiating new collaborations, both parties need to discuss the goals and expectations of the collaboration. Moreover, both parties need to spend time on the collaboration and to learn about the disciplinary areas of all of the partners. Additionally, both parties need to clearly understand the individual benefits to be gained from the collaboration. This is important to achieve buy-in for the collaboration as well as for career advancement.

5. **Education:** Dr. Tinsley Oden shared with workshop attendees a successful model of immersive interdisciplinary graduate education currently in place at the Institute for Computational Engineering and Sciences at the University of Texas at Austin. Specifically, beginning graduate students in the program must take a specific set of courses spanning advanced mathematics and computer science, as well as physics, biology, and others. This introduces students to concepts in relevant disciplines. It also gives them a common language with which to discuss problems. Lastly, it gives them a basic foundation on which to then pursue interdisciplinary computational research.

B. Challenges in the Field:

6. **Data:** Many research challenges are found within the area of data, especially questions on storing, moving, interpreting, visualizing, and quantifying uncertainties in large data sets.

7. **People:** As Dr. Miron Livny posited, Cyberinfrastructure = Hardware + Software + People. The community needs to shape and define the "data scientist" of the coming decade and to develop the educational programs focused on computational- and data-enabled science and engineering (CDS&E).

8. **Engineering vs. Science:** We discussed how cyberinfrastructure is a science and engineering discipline and how investments can help to fuel a virtuous cycle of science → engineering → application/use → new discoveries.

9. **Sustainability:** We discussed the inherent problem of sustainability for software and data created as a part of NSF-funded research, and the need to devise new models for sustainability that will uphold the academic values of open research and inquiry as well as the needs of the marketplace.

10. **Cyberinfrastructure Framework:** What common elements of cyberinfrastructure can we identify across the range of cyberinfrastructure sytsems that can lead to a definition of a common framework for cyberinfrastructure?

Trends and Observations from the Workshop

We summarize trends and observations of the workshop as indicated by NSF CAREER attendees in workshop discussions as well as informal conversations with workshop organizers and staff.

- Faculty appreciated the power and potential of interdisciplinary research that involves both pathbreaking work within a single discipline and significant expertise and work from outside of their primary discipline.
- Attendees were inspired by the talks from the senior keynote speakers on interdisciplinary research topics.
- Faculty are very interested in open source software, dissemination of software products and data produced by the faculty, and new ways of gaining promotion and tenure credit for software and data products produced and adopted by the community.
- Faculty noted that many issues surround promotion and tenure for interdisciplinary activities. Many attendees felt that they did not fit cleanly within a department, for example. Moreover, there is often concern about how to best determine the most appropriate NSF division to which to submit an interdisciplinary proposal.

- Attendees indicated that although it was helpful to learn more about the NSF at the workshop, they would also like to see program directors from NIH invited to give presentations.
- Faculty indicated that they liked the single track of the workshop.
- Faculty also liked the pacing of the workshop with generous time made available for discussion. Small group breakouts and discussions also provided opportunities for in-depth discussion as well as networking opportunities.
- Faculty would like to have the option to give a lightning talk on their research. They thought this would help them get to understand NSF CAREER attendee research more easily than did the poster session, for example.

Workshop Attendees

Twenty-five attendees and five keynote speakers attended the workshop, from twenty-three institutions (Fig. 1). Most attendees were funded at least in part through the Office of

Cyberinfrastructure, but several held awards partially funded through other directorates, including Mathematical and Physical Sciences, Engineering, Education and Human Resources, Biological Sciences, and Computer and Information Science and Engineering (Fig. 2). Seven attendees were funded entirely from directorates outside the OCI.

Attendee Selection Process:

Faculty who received NSF CAREER Awards from OCI (including through co-funding with another NSF directorate or division) were invited to attend the workshop. The list of NSF OCI CAREER faculty was provided to us by Gabrielle Allen. Many of these faculty accepted our invitation to attend the workshop. However, there were still several seats available in the workshop due to travel conflicts, etc. Thus, we made a decision to invite several NSF CAREER faculty with related research projects (e.g., faculty whose NSF CAREER projects involve computation and an application). Inviting a broader range of faculty also served another purpose which was having a larger array of faculty in attendance to discuss a wider range of research and education collaborations. Gabrielle sought recommendations of NSF CAREER Awardees whose projects focused on related research from NSF program directors in relevant divisions or directorates (e.g., MPS, ENG, CISE, BIO, etc.). The members of the organizing committee were also encouraged to recommend several NSF CAREER faculty with related research projects. The organizing committee came up with a list of several names using the NSF Fastlane system to perform searches by research topic; these names were then submitted to Gabrielle for approval. Invitations were then sent to the list of NSF CAREER faculty with related research projects sent to us by Gabrielle. We received enough acceptances of our invitations from faculty in this category to fill all of the seats in our workshop.

Figure 1. Map of the United States showing locations of the various attendee home institutions at the CyberBridges Workshop.

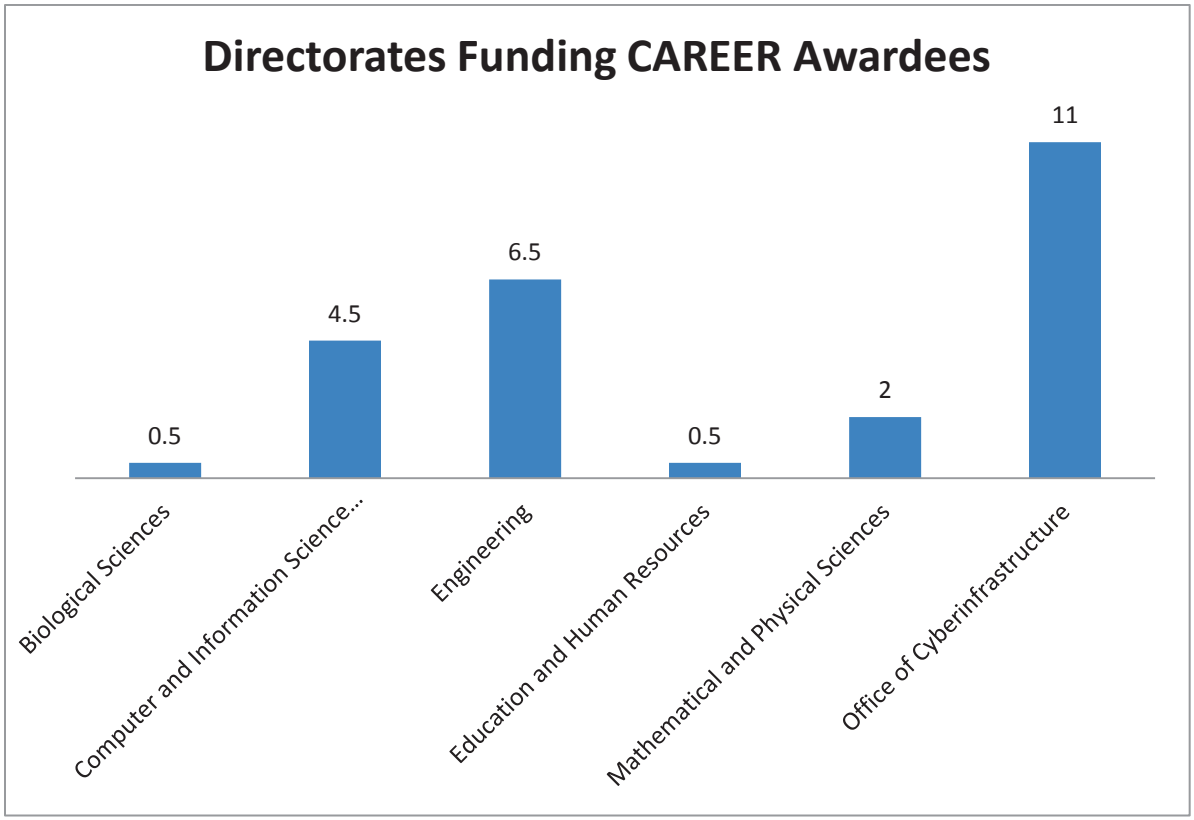


Figure 2. Distribution of the directorates funding the attendees of the CyberBridges Workshop. CAREER Awards funded through dual directorates are split evenly between the relevant offices and directorates.

Workshop Themes

The workshop focused on five thematic areas that reflect the spectrum of research and education activities in which the Office of Cyberinfrastructure (OCI) is involved, and they encompass the types of computational- and data-enabled science and engineering (CDS&E) in which other NSF directorates are engaged.

1. Grand Challenges in CDS&E and Cyberinfrastructure

The first thematic area of the workshop focused on *predictive science*, which seeks to use computational and data analysis to model and predict the future behavior of physical and biological systems. Dr. Tinsley Oden provided a keynote talk in this area and led discussion among workshop participants. The fundamental elements of predictive science include (1) model validation, which focuses on assessing the accuracy of the prediction; (2) verification, which determines the degree of correspondence between computational and mathematical models; (3) calibration, which is the identification and tuning of model parameters to bring computational model predictions into agreement with experimental measurements; and (4) identifying the target outputs and the objectives/goals of a model. Examples of predictive science are weather and climate prediction, drug design, nanomanufacturing, and individualized patient specific modeling of the human body. The issues in this area described by Dr. Oden and discussed in the workshop include (1) the need for an education program focused on providing interdisciplinary training with strong emphasis on core math/science/computation/applications; (2) the need for best practices to facilitate data and software sharing; (3) the problems involved in managing and using large amounts of data and how to link computation with experiments; (4) the inherent cultural issues involved in developing and sustaining interdisciplinary collaboration, including organizational structures and how to convince others in your research areas that interdisciplinary work is adequately rigorous and meaningful; (5) the need to educate citizens about sciences; and (6) the need to expand the definition of publication to include data and software contributions.

2. Data-enabled Science and Engineering

The second thematic area of the workshop focused on *data-enabled science and engineering*, which involves the development of algorithms necessary to process and interpret data generated from experiments, simulations, models, and observations in science and engineering. Dr. Frederica Darema provided a keynote talk in this area and led discussion among workshop participants. Portions of keynote talks given by

Ed Seidel and Miron Livny, along with the ensuing discussion among workshop participants, were also relevant to this thematic area. The fundamental issues in data-enabled science and engineering include: (1) developing efficient algorithms to address problems in Big Data, i.e., applications that collect terabytes of data (such as the Sloan Digital Sky Survey and the Large Synoptic Survey Telescope); (2) combining data and computation into effective algorithms; (3) the need to visualize large datasets; (4) the need to develop a unified model of high-end computing (high-performance and high-throughput computing, grid computing, and cloud computing) that will make it unnecessary to 'reinvent the wheel' and to facilitate the integration of real-time data collection, analysis, and decision making; (5) the need for open-source software that can be sustainably shared; (6) the need for industry university partnerships to address these issues, and (7) how to fund research, as well as the software development and data dissemination activities needed across communities. The National Science Foundation has fostered two initiatives that seek to address the challenges in data-enabled science and engineering: the cross-disciplinary Computational and Data-Enabled Science & Engineering Program; and work to address selected Grand Challenges (which also address other large-scale challenges as described in the NSF Advisory Committee on Cyberinfrastructure reports). Dr. Darema also discussed her Dynamic Data-Driven Applications Systems (DDDAS) initiative, which is an example of a framework that synthesizes data and computation to create a symbiotic feedback control system. DDDAS includes the ability of the system to dynamically incorporate additional data as it becomes available during the application, and the ability to dynamically steer the application process. One example of DDDAS application involved network systems science and how to understand the brain and the mind.

3. Scientific Visualization

The process of interpreting and distilling knowledge from data is greatly facilitated by converting these data into *visualizations* to allow users to quickly understand data. The tremendous volumes of data that need to be interpreted are driving the discovery and development of new visualization technologies. These technologies are used in diverse application areas such as biology, physics, and earthquake engineering. Emerging areas of visualization research described by Dr. Chris Johnson included 3D uncertainty quantification, algorithmic analysis of images, very-high-resolution visualization, and the statistics of shape, connectivity, and function. The discovery and development of these new approaches and technologies are strongly driven by a pressing need to solve real-world problems. An example described by Dr. Johnson is the need to quantify spatial "error bars" in 3D of the surface boundaries of brain tumors collected through Magnetic Resonance (MR) imaging. Effective modern

radiation treatment using a "cyber knife" depends on the ability to accurately identify the location and boundaries of tumors, which can be improved through the use of 3D uncertainty quantification computed from MR data.

4. Computational-Enabled Science and Engineering

The fourth thematic area of the workshop was *computational-enabled science and engineering*. Computational science is the development of numerical algorithms for the solution of mathematical problems that arise in scientific applications. The numerical algorithms that are developed typically have a close tie to the scientific application of interest. Dr. Tinsley Oden's keynote talk addressed this theme (along with Grand Challenges in CDS&E), and he led the ensuing discussion among workshop participants. A portion of the talk by Dr. Ed Seidel was also related to this theme. One important topic that was emphasized within this theme was a paradigm shift which is occurring in science, engineering, and education. Research is becoming more interdisciplinary, multidisciplinary, and transdisciplinary. Grand challenge communities have developed to tackle very difficult societal problems. Examples of community projects include the community iPlant toolkit, the Laser Interferometer Gravitational-Wave Observatory (LIGO), and the Great Lakes Consortium for Petascale Computation Blue Waters system. These challenging applications require researchers to work at all scales and to make effective use of large-scale computing resources. CDS&E was presented as an emerging research community. Within this community, there is an emphasis on the increasing importance of data that should be coupled with numerical algorithms developed in computational science and engineering programs. The need for researchers and practitioners who are aware of and skilled in the development and use of CDS&E methodologies reveals the important need to develop new education programs that are more interdisciplinary, multidisciplinary, and transdisciplinary.

5. High-Performance and High-Throughput Computing

The fifth thematic area of the workshop was high-performance (HPC) and high-throughput computing (HTC). This area focuses on the development and use of HPC and HTC systems for solving problems in science and engineering that could not have been solved without the use of these systems. Dr. Miron Livny presented a keynote talk on the area and led a subsequent discussion with workshop participants. A key question that drives this area was posed by Dr. Livny: should the development and building of cyberinfrastructure in support of research be considered a science, or is it engineering? Real-world problems that require the use of HPC and HTC technologies can help to expose difficult and "real" computer science problems that

need innovative and transformative solutions. By working on these real and difficult problems, computer scientists have the best chances of making an impact on the field and to conduct meaningful and relevant science. Through a sustained effort to address these difficult and real problems, computer-science researchers can build successful careers as academics over the long run.

Invited Speakers and Panelists from the National Science Foundation

Dr. Alan Blatecky's talk focused on data-enabled science and engineering (as described above), the research challenges in this area, and the efforts of the National Science Foundation in this area. According to Dr. Blatecky, the goal of data-enabled science and engineering research is to develop algorithms to extract knowledge from very large datasets to address grand-challenge community problems, for example. By 2020, we will have petabytes to exabytes of data from various applications and will need algorithms for processing these data in a distributed manner at all times (as opposed to distributing the data at the end). Moreover, the useful lifetime of data should be significantly increased from where it is today.

The NSF now requires data management plans for proposals submitted to NSF – the open access of data and software is the goal unless a compelling reason forbids sharing them. The NSF has developed several data-oriented solicitations that include SSE, SSI, SSII, DIBBS, Earth Cube, Big Data, CDS&E, CC-NIE, SEES, and CIF21. The long-term storage of data and software is also a concern. The current practice in the research community is that publications are the mechanism of storage of research results. A shift is needed in the thinking of tenure and promotion committees to increase their emphasis on the creation and dissemination of community data and software products.

Dr. Gabrielle Allen's talk focused on an introduction to the Office of Cyberinfrastructure (OCI), cyberinfrastructure (CI) resources, the Grand Challenges Task Force foci, and funding for students and postdoctoral associates. The focus of OCI is on advanced computing infrastructure, software, data, networks and security, and learning and workforce development in these areas. CI resources include XSEDE (an integrated platform for supercomputers, data, and computational tools), the Open Science Grid, international networks, computing technologies, mathematical algorithms, scientific challenges, organizational technologies, and educational activities. The Grand Challenges Task Force focuses on the areas of computational- and data-enabled science and engineering, high-performance computing, data, visualization, applied and computational math, computer science, and core science and engineering disciplines. Relevant NSF solicitations in OCI are in Strategic Technologies for Cyberinfrastructure (STCI) and the Computational-and Data-Enabled Science and Engineering (CDS&E) areas. Within OCI, prospective postdoctoral associates may apply for CI-TRACS

fellowships; REU supplements to NSF projects are available for undergraduate student researchers; IGERT funding is available for graduate students.

Dr. Daniel Katz's talk focused on software efforts at OCI. OCI is interested in building the software ecosystem through new capabilities while enabling transformative, interdisciplinary, and collaborative science and engineering research and education. OCI is interested in research that will benefit multiple disciplines. OCI is building up the software ecosystem through software solicitations that start with smaller projects (i.e., software elements) that build into larger projects (i.e., software frameworks) that transition into community projects (i.e., software institutes) that emphasize reuse of software and use by large groups of researchers. Such projects should budget a certain number of full-time equivalents to keep things going.

Dr. Dane Skow presented an overview of data-oriented solicitations from OCI (BIGDATA and DIBBS) and described the EarthCube initiative. He described some of the big data problems facing the community today, which include multiple sources of data, the need to solve data-oriented interdisciplinary problems, and the need to provide a reliable data repository to avoid "data stockpiling," which is becoming a problem within research laboratories. For the future, Dr. Skow described the need for a "data fabric" to be shared within and across disciplines to provide a stable and reliable platform for research data produced by the community.

Dr. Eduardo Misawa described the CIF21 program and the importance of this program to the Engineering communities. He described the need to maintain a balance in investments between computational and data-enabled research, and the available programs for CDS&E research that include CDS&E-ENG, GOALI, and cross-directorate special calls for proposals.

Dr. Jean Cottam Allen described cyberinfrastructure activities in the Mathematics and Physical Sciences Directorate that are focused on addressing the cyber-related needs of MPS science communities. She described the common goals for cyberinfrastructure, which include (1) the need for long-term data preservation and reuse, data accessibility, and the need for new approaches for extracting information from data; (2) support for the development of new software tools and algorithms; (3) support for disciplinary scientists to employ advanced architectures; (4) network security along with hardware and software bridging that go beyond the Internet protocols; (5) workforce development; and (6) the need for Grand Challenge projects that address significant and transformative science that can be addressed only through the use of cyberinfrastructure. Dr. Allen described the current MPS cyberinfrastructure activities that include BIGDATA, DIBBS, Software Infrastructure for Sustained Innovation, and Physics at the Information Frontier – Computational Physics.

Dr. Peter McCartney described some of the cyberinfrastructure-related activities in the BIO directorate, which included the Protein Data Bank (PDB) and the BIO Synthesis Center.

Dr. Barry Schneider described the XSEDE program, and the process researchers can follow to request access to XSEDE resources. He described efforts of the XSEDE program to host research data that are intended to be shared with the community. Dr. Schneider described some of the challenges involved in developing and operating a shared-data infrastructure, which include sustainable funding models that go beyond the initial funding needed to create data infrastructures.

Attendee Feedback Survey

To collect attendee feedback, we conducted an anonymous Qualtrics survey that was emailed to workshop attendees after the workshop. The purpose of this survey was to collect feedback from attendees about what went well and what did not go well, and to ask for suggestions for improvement for the workshop. We received 13 responses to the survey (see the Appendix A for the complete text of the survey and responses).

Based on survey results, participants strongly agreed that the five thematic areas of the workshop included their areas of research and education, and that the disciplinary areas of workshop attendees were sufficiently broad to facilitate interdisciplinary engagements (Table 3). Participants also strongly agreed that that workshop format (talks followed by discussion) was useful and engaging, the keynote talks were informative and interesting, the poster session was useful and engaging, and that the workshop was helpful to learn more about the NSF and available funding opportunities. Participants slightly less strongly agreed (4.2) that there were sufficient opportunities for networking and collaboration.

In terms of participants' perceptions about future participants, results indicate strong agreement (4.2) that the workshop should include CAREER awardees beyond OCI, but only low neutral agreement (3.1) that the workshop should include attendees outside the CAREER program.

Table 3. Participant responses to the workshop survey. Responses are mean and standard deviation (S.D.). Survey participants answered questions on a Likert Scale with the following numeric assignment. Strongly Disagree (SD = 1), Disagree (D = 2), Neither Agree nor Disagree (AD = 3), Agree (A = 4), and Strongly Agree (SA = 5).

Question	Mean	S.D.
The five focus areas of the workshop included my area of research and education.	4.5	0.52
The disciplinary areas of workshop attendees were sufficiently broad to facilitate interdisciplinary engagement.	4.5	0.52
The workshop format (keynote talks followed by discussion) was useful and engaging.	4.5	0.52
The talks were relative, informative, and interesting.	4.4	0.51
The poster session was useful and engaging.	4.5	0.66
There were sufficient opportunities for networking and collaboration.	4.2	0.6
The hotel accommodations, meeting space, and meals were adequate.	4.8	0.44
The workshop was helpful in learning more about the NSF and available funding opportunities.	4.7	0.48
The workshop should include CAREER awardees beyond OCI.	4.2	0.69
The workshop should include attendees from outside the NSF CAREER program.	3.1	1.19

Participants felt that the workshop length and the number of attendees were about right. Also, 85% of the survey responses indicated interest in attending the workshop even if full travel reimbursement has not provided.

Additional questions provided an open form that allowed participants to provide written feedback on additional topics that could be covered in future workshops, suggestions for improvements, and suggestions for future keynote speakers. Suggestions for additional topics included access to HPC resources, advanced networking and big data, more specificity on ‘grand challenges’ (scientific questions and funding directions), computational issues in stochastic

simulations, climate change modeling, and systems engineering, and algorithms for medicine, energy, data mining, and other 'hot topics'.

Several participants suggested that it would be useful to have short presentations (5-10 minutes) from all of the attendees to highlight their research. Two respondents asked for more social and networking time, with less focus on organized talks, and one suggested that the workshop should also include program directors from other agencies (eg NIH, DoE, and DoD) that fund this work.

Lessons Learned From the Workshop

We learned several lessons by organizing the first NSF CyberBridges Workshop. Below we focus on workshop aspects that should be changed for future workshops along with set of specific observations from the workshop.

Aspects that should be changed for future workshops

A more specific and well-defined process for inviting NSF CAREER Awardees to attend the workshop.

Well-structured lightning talks should be added to provide an overview of attendee research and education activities.

A brief pre-workshop biosketch of attendee research interests should be made available one week before the workshop.

We should devise a more formal and structured approach for managing breakout sessions after each keynote talk and for collecting feedback from breakout groups. We should ask speakers to select one approach for breakout groups from a limited menu of options (e.g., 1 group vs. 5 breakout groups).

Co-funded NSF CAREER awardees should be encouraged to invite their program manager outside of the Office of Cyberinfrastructure to view their poster during the poster session.

We should increase the focus on education activities.

We should increase the disciplinary diversity of the keynote speakers (i.e., include a speaker from Mathematics).

Observations from the Workshop

The workshop provided many opportunities for discussions among attendees and speakers. We received many comments from the attendees. In this section, we distill these comments into a set of observations that reflect the thoughts of attendees.

Observation 1. The attendees felt that there is a lack in the number of interdisciplinary solicitations that bridge the domain sciences, computer science, and cyberinfrastructure.

Observation 2. The attendees suggested that the CAREER proposal review criteria could be expanded to encourage the development and dissemination of open-source software and data as part of the proposed project.

Observation 3. The attendees described an unmet need for a central cyberinfrastructure that could be used by CAREER awardees to disseminate open-source software and data (an example of this type of system is the NEEShub)

Observation 4. To assist early-career researchers in developing their research program and in seeking federal sponsorship for their research, many attendees described a need for clear guidelines to CAREER awardees on the science and engineering areas covered by NSF programs and solicitations and when researchers should approach other agencies, such as NIH and DoE.

Observation 5. The NSF CyberBridges Workshop provided a productive venue for CAREER awardees to meet, network, and explore integrative and joint research and education activities, and to allow senior researchers to provide career advice to young researchers.

Observation 6. There is a need to create more core solicitation programs with well-defined submission schedules to allow researchers to better plan in advance to submit proposals to a specific program, rather than the need to respond more quickly to singular solicitations that are not part of a recurring program.

The recent move of OCI into CISE brings up several questions.

- The details of the realignment need to be communicated as soon as possible to the community.
- Researchers working on OCI CAREER proposals will need guidance as soon as possible if there are to be changes to the program.

Acknowledgements

Our efforts in planning and conducting the CyberBridges workshop and the development of this report were supported by Natasha Nikolaidis (Purdue), Amy Hasan (Penn State), Jenny

Latchford (Penn State), Stephanie Schmidt (Purdue), Lena Pipenberg (Penn State), and Debbie Miethke (Purdue).

Appendix A

Detailed Survey Results

The complete text and responses of the survey sent out to attendees are described below.

In the first section, survey participants answered questions on a Likert Scale with the following numeric assignment. Strongly Disagree (SD = 1), Disagree (D = 2), Neither Agree nor Disagree (AD = 3), Agree (A = 4), and Strongly Agree (SA = 5).

1. The five focus areas of the workshop (Grand Challenges, Data, Visualization, Computational Science, and High Performance Computing) included my area of research and education).
   Results: Mean Value: 4.5. Responses: (6) A, (7) SA

2. The disciplinary areas of workshop attendees were sufficiently broad to facilitate interdisciplinary engagement.
   Results: Mean Value: 4.5. Responses (7) A, (6) SA

3. The workshop format (keynote talks followed by discussion) was useful and engaging.
   Results: Mean Value: 4.5. Responses (6) A, (7) SA

4. The talks were relative, informative, and interesting.
   Results: Mean Value: 4.4. Reponses (8) A, (5) SA

5. The poster session was useful and engaging.
   Results: Mean Value: 4.5. Reponses (1) AD, (5) A, (7) SA

6. There were sufficient opportunities for networking and collaboration.
   Results: Mean Value: 4.2. Responses (1) AD, (8) A, (4) SA

7. The hotel accommodations, meeting space, and meals were adequate.
   Results: Mean Value: 4.8. Responses (3) A, (10) SA

8. The workshop was helpful in learning more about the NSF and available funding opportunities.
   Results: Mean Value: 4.7. Responses (4) A, 9 (SA)

9. The workshop should include CAREER awardees beyond OCI.
   Results: Mean Value: 4.2. Responses (2) AD, (7) A, (4) SA

10. The workshop should include attendees from outside the NSF CAREER program.
    Results: Mean Value: 3.1. Responses (5) D, (5) AD, 3 (SA)

Next, the survey asked participants to rate the following questions with a response from the options Too short/too few (1 = S); About right (2 = AR); and Too long/too many (3 = L).

11. Length of the workshop
    Results: Mean Value: 1.9. Responses (2) S, (11) AR

12. Number of attendees
    Results: Mean Value: 1.9. Responses (1) S, (12) AR

The next two questions asked for a Yes/No response:

13. Would you be interested in attending the workshop in the future without full travel reimbursement?
    Results: Yes 85%, No 15%

14. Is a component of your CAREER award funded from OCI?
    Results: Yes 54%, No 46%

The final questions provided an open form to allow participants to provide written feedback:

15. Are there any new broad areas or topics that you would like to see covered in a follow on workshop?
    Responses:
    a. Computational issues on: Stochastic simulations, modeling of climate change, systems engineering
    b. More specificity on 'grand challenges' -- scientific questions and funding directions
    c. computational science algorithms for medicine, energy, data mining, and other "hot topics"
    d. Application-specific cyberinfrastructure topics
    e. Advanced networking, big data
    f. More focus on collaborative discussions, less on talks
    g. Access to HPC resources

16. What changes or improvements could we make to the workshop in the future?

Responses:

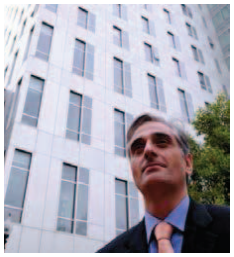
a. Possibly ask attendees to introduce their research by delivering a very short (5-minute) presentation
b. Brief (10 min) oral research highlights from the participants
c. I would like to see the number of attendees grow somewhat. I would also like to see program directors invited from some of the federal agencies which fund this work invited (e.g., NIH, DoE, and DoD). I would also love for us to have a brainstorming session in which we discover what is "missing" in the application domains and think up "computational solutions".
d. More social time
e. 5 minutes presentation for attendees in sessions (parallel or non-parallel), which could be shorter or longer

Appendix B

Attendee Biographies and Photos

Lorena Barba

Lorena A. Barba is an Assistant Professor of Mechanical Engineering at Boston University. She works in computational fluid dynamics, especially immersed boundary methods and particle methods for fluid simulation; fundamental and applied aspects of fluid dynamics, especially flows dominated by vorticity dynamics; the fast multipole method and applications; and scientific computing on GPU architecture. http://barbagroup.bu.edu/



Luca Caracoglia

Luca Caracoglia is an Associate Professor in the Department of Civil and Environmental Engineering of Northeastern University, Boston, Massachusetts, USA. His interest are in structural dynamics, random vibrations, passive structural control, wind engineering, fluid-structure interaction of civil engineering structures, linear and nonlinear cable network dynamics, and wind-based energy harvesting systems. Luca Caracoglia received the NSF-CAREER Award from the CMMI Division in 2009. http://www1.coe.neu.edu/~lucac/default.htm

Diego Donzis

Diego Donzis is an Assistant Professor in the Department of Aerospace Engineering at Texas A&M University. He is interested in large-scale computing, fluid mechanics, and turbulence and turbulent mixing in incompressible and compressible flows. He obtained his PhD at Georgia Tech and worked in the University of Maryland and the International Center for Theoretical Physics (Italy) before joining the faculty at Texas A&M. http://aero.tamu.edu/people/faculty/?id=529

Gabriel Dos Reis

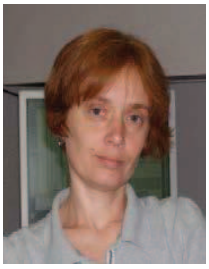
Gabriel Dos Reis is an Assistant Professor in the Department of Computer Science and Engineering at Texas A&M University. He works on theoretical and practical aspects of computing, and his research interests include software systems and tools, programming methodologies, tools and applications of formal methods, mathematical aspects of software construction, and symbolic mathematics. http://parasol.tamu.edu/~gdr/

Maria Emelianenko

Maria Emelianenko was born in Dubna, Russia. She received an MS in Applied Mathematics from Moscow State University and a PhD in Mathematics from Pennsylvania State University in 2005. She spent two years as a Research Associate at the Carnegie Mellon Center for Nonlinear Analysis before joining the George Mason University faculty in 2007, where she is now an Associate Professor. She is currently working on problems arising on the interface between mathematics, physics, biology, and engineering. http://math.gmu.edu/~memelian/

Baskar Ganapathysubramanian

Baskar Ganapathysubramanian is an Assistant Professor of Mechanical Engineering and Electrical and Computer Engineering at Iowa State University. His research interests are in stochastic analysis, multiscale modeling, and design of materials and processes using computational techniques. Dr. Ganapathysubramanian completed his PhD and MS from Cornell University and holds a BS degree from the Indian Institute of Technology – Madras. http://www3.me.iastate.edu/bglab/



Sophya Grashchuk

Sophya Grashchuk is a theoretical and computational chemist at the University of South Carolina since 2008. She received the NSF CAREER award in 2011. Sophya's research interests include development of trajectory-based approximate quantum molecular dynamics and studies of quantum nuclear effects on reactivity in complex molecular environments. http://www.chem.sc.edu/people/facultyStaffDetails.asp?SID=83

Thomas Hacker

Tom Hacker is an Associate Professor of Computer and Information Technology at Purdue University. Dr. Hacker's research interests center around high-performance computing and networking on the operating system and middleware layers. Recently his research has focused on cloud computing, cyberinfrastructure, scientific workflows, and data-oriented infrastructure. Dr. Hacker is also coleader for Information Technology for the Network for Earthquake Engineering Simulation (NEES), which brings together researchers from fourteen universities across the country to share innovations in earthquake research and engineering. http://www2.tech.purdue.edu/cpt/SelfStudy/CPTFacultyVitas/FacultyStaff/DisplayStaffMember.asp?member=tjhacker

Richard Hennig

Professor Hennig received his Diploma in Physics at the University of Gottingen in 1997 and his PhD in Physics from Washington University in St. Louis in 2000. After working as a postdoctoral researcher and research scientist at Ohio State University, he joined the faculty of the Department of Materials Science and Engineering at Cornell University in 2006.
http://theory.mse.cornell.edu

Kapil Khandelwal

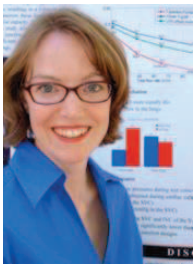
Kapil Khandelwal is an Assistant Professor in the Department of Civil Engineering and Geological Sciences at Notre Dame University. He received his PhD in Civil and Structural Engineering from the University of Michigan in 2008. He is currently interested in how infrastructural systems behave under extreme load conditions. Much of his research is focused on the computational and theoretical aspects of structural engineering, with particular emphasis on multiscale computational simulations, including finite element methods, structural optimization, constitutive modeling, damage-plasticity formulations, nonlinear solution strategies, and advanced visualization techniques. http://www.nd.edu/~kkhandel/

Xiaolin Li

Dr. Xiaolin (Andy) Li is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Florida. His research interests include cloud and high-performance computing and cyber-physical systems. He is the director of the Scalable Software Systems Laboratory (S3Lab), designing large-scale software systems and tools for enabling high-impact applications in science, engineering, and health care. He received the NSF CAREER Award in 2010.
http://www.andyli.ece.ufl.edu/

Laurence Loewe

Dr. Laurence Loewe is an Assistant Professor at the University of Wisconsin-Madison. His research focuses on understanding the molecular, genetic and ecological basis for adaptive evolution and extinction in natural populations and its implications. He is interested in bridging the gap between simple analytically understandable mathematical models and biological reality by building rigorous simulation models to answer various evolutionary questions.
http://evolution.ws/people/loewe



Alison Marsden

Alison Marsden is an Assistant Professor of Mechanical and Aerospace Engineering at UCSD. She holds a BS from Princeton and a PhD from Stanford. She received a BWF CASI award in 2007 and an NSF CAREER award in 2012. Her research focuses on cardiovascular simulations, optimization, and congenital heart disease.
http://maresearch.ucsd.edu/marsden/AMarsden/Home.html

Christian Ott

Christian Ott is a computational astrophysicist studying supernova explosions, neutron stars, and black holes. He received his Diploma in Physics from the University of Heidelberg in 2003 and his PhD from the Albert Einstein Institute and the University of Potsdam in 2007. Before being appointed as Assistant Professor at California Institute of Technology (Caltech) in 2009, he was a postdoctoral fellow at the University of Arizona and at Caltech.
http://www.tapir.caltech.edu/~cott

Dario Pompili

After graduating from the University of Rome "La Sapienza" and Georgia Tech, in 2007 Dr. Pompili joined Rutgers/ECE as an Assistant Professor, where he co-directs the NSF CAC Center. In 2011, he received the NSF CAREER on underwater communications. In 2012, he received the ONR YIP on mobile computing and the DARPA YFA on real-time advanced neuroscience. http://www.ece.rutgers.edu/~pompili/

Iaon Raicu

Ioan Raicu is an assistant professor in CS at Illinois Institute of Technology, and a guest research faculty in MCS at Argonne National Laboratory. He is also the founder and director of the Data-Intensive Distributed Systems Laboratory. My research work and interests are in the area of distributed systems. http://www.cs.iit.edu/~iraicu/

Pradeep Ravikumar

Pradeep Ravikumar received his BTech in Computer Science and Engineering from the Indian Institute of Technology, Bombay, and his PhD in Machine Learning from the School of Computer Science at Carnegie Mellon University. He then became a postdoctoral scholar at the Department of Statistics at the University of California at Berkeley. He is now an Assistant Professor in the Department of Computer Science at the University of Texas at Austin. He is also affiliated with the Division of Statistics and Scientific Computation and the Institute for Computational Engineering and Sciences at UT Austin. His thesis has received an honorable mention in the ACM SIGKDD Dissertation competition and the CMU School of Computer Science Distinguished Dissertation Award.
http://www.cs.utexas.edu/~pradeepr

Pratim Sengupta

Pratim Sengupta is an Assistant Professor of Learning Sciences and Science Education at Vanderbilt University, where he also directs the Mind, Matter, and Media Lab. His research focuses on designing agent-based programming languages for K12 science education. http://www.m3lab.org/people/pratim

Suzanne Shontz

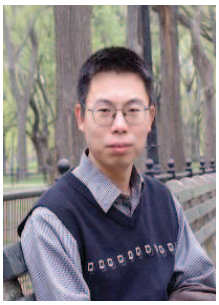
Suzanne Shontz is an Assistant Professor in the Department of Computer Science and Engineering at The Pennsylvania State University. Suzanne's research is in parallel scientific computing and focuses on the development of meshing and numerical optimization algorithms and their applications. In 2011, Suzanne received an NSF CAREER Award.
http://www.cse.psu.edu/~shontz

Andres Tejada-Martinez

Andres Tejada-Martinez is an Assistant Professor in Civil and Environmental Engineering at the University of South Florida. Tejada-Martinez has received an NSF CAREER Award and various other NSF collaborative research awards for his work in large-eddy simulations of turbulent mixing in shallow-shelf coastal regions and in the upper-ocean mixed layer.
http://www.eng.usf.edu/~aetejada/journal.html

Xavier Trioche

Xaiver Trioche (PhD, 2002, Computer Science, University of Kaiserslautern, Germany) is Assistant Professor of Computer Science at Purdue University. His main research interests include interactive visualization, structural data analysis at scale, and computational steering, with primary applications in fluid dynamics, medical image analysis, high-energy physics, and granular materials. http://www.cs.purdue.edu/homes/nmt



Liqiang Wang

Liqiang (Eric) Wang has been an Associate Professor (2012-present) and an Assistant Professor (2006-2012) in the Department of Computer Science at the University of Wyoming. He received his PhD in Computer Science from Stony Brook University in 2006. His research interest is the design and analysis of parallel systems. For analysis, he is working mainly on concurrency/security-related error detection. For design, he is currently working on data-intensive parallel computing on multicore CPU, GPU, and cloud-computing platforms. He is also interested in integrating parallel computing with scientific work flows. His research has been supported by NSF, ONR, NASA, and the University of Wyoming. He received the NSF CAREER Award in 2011. http://www.cs.uwyo.edu/~lwang7/

Clint Whaley

R. Clint Whaley is an Assistant Professor in Computer Science at the University of Texas, San Antonio. He is the author of the ATLAS linear algebra tuning platform and the empirical compilation framework iFKO. http://www.cs.utsa.edu/~whaley/

Xiong Yu

Xiong (Bill) Yu, an Associate Professor of Civil Engineering at Case Western Reserve University, also holds a courtesy appointment in Electrical Engineering and Computer Science and also in Mechanical Engineering. His research includes smart and functional materials, intelligent structure, sensor innovations, and computer-assisted engineering. He received his NSF CAREER award in 2009. http://filer.case.edu/xxy21/Index.html

Jessica Zhang

Jessica Zhang is an Associate Professor in Mechanical Engineering at Carnegie Mellon University with a courtesy appointment in Biomedical Engineering. Her research interests include computational geometry, mesh generation, computer graphics, visualization, finite element method, isogeometric analysis, and their applications to computational biomedicine and engineering. http://www.andrew.cmu.edu/user/jessicaz/

Appendix C

Poster Session

The poster session provided a forum for attendees to present work from their CAREER projects and facilitated connections and collaborations among researchers. All of the workshop participants presented posters. Poster topics included topics that involved high-performance computing, parallel computing and distributed storage systems, fault tolerance and error detection, and data visualization. Moreover, several researchers presented applied work where high-performance computing was being used to develop models and algorithms to address questions in various fields, such as astrophysics, materials and structural engineering, fluid dynamics, biology, and medicine. The posters presented are listed in Table 1:

Table 1. Posters presented at the CyberBridges conference.

Last Name	First Name	Institution	Poster Title
Barba	Lorena	Boston University	Scalable algorithms for extreme computing on heterogeneous hardware.
Caracoglia	Luca	Northeastern University	Dynamic performance of tall buildings against wind hazards: recent results on a new simulation framework.
Donzis	Diego	Texas A&M University	Discoveries in compressible turbulence and turbulent mixing through petascale simulation and analysis
Dos Reis	Gabriel	Texas A&M University	The role of programming languages and tools in dependable systems.
Emelianenko	Maria	George Mason University	Multiscale computational modeling of coarsening in materials.
Ganapathysubramanian	Baskar	Iowa State University	A cyber-enabled approach to unraveling process-structure-property relationships in organic electronics.
Garashchuk	Sophya	University of South Carolina	Modeling quantum effects of moving nuclei in biological systems, carbon materials, and other molecular environments.
Hacker	Thomas	Purdue University	Developing new approaches for fault avoidance and fault tolerance for large-scale HPC systems.
Henning	Richard	Cornell University	Computational tools for the discovery of materials structures and interface properties.
Khandelwal	Kapil	Notre Dame University	Computational challenges in structural engineering: from simulation to practice.

Table 1, cont'd

Li	Xiaolin (Andy)	University of Florida	Scalable adaptive runtime management algorithms and tool kit.
Loewe	Laurence	University of Wisconsin-Madison	Modeling made easy.
Marsden	Alison	University of California, San Diego	Optimization and parameterization for multiscale cardiovascular flow simulations using high- performance computing.
Ott	Christian	California Institute of Technology	Cyberinfrastructure for computational relativistic astrophysics.
Pompili	Dario	Rutgers University	Investigating fundamental problems for underwater multimedia communication with application to ocean exploration.
Raicu	Ioan	Illinois Institute of Technology	Building blocks for scalable distributed storage systems.
Ravikuar	Pradeep	University of Texas at Austin	Component-based models: a unified "neat" framework for statistical machine learning.
Sengupta	Pratim	Vanderbilt University	CORE-Science: computational reconstruction of K12 science.
Shontz	Suzanne	The Pennsylvania State University	Parallel dynamic meshing algorithms for patient-specific medical interventions.
Tejada-Martinez	Andres	University of South Florida	Disruption of bottom-log-layer in LES of full depth Langmuir circulation.
Trioche	Xavier	Purdue University	A scalable visualization infrastructure to empower the scientific community.
Wang	Liqiang	University of Wyoming	Towards scalable error for detection for parallel software systems on emerging computing platforms.
Whaley	Clint	University of Texas at San Antonio	Empirical tuning for extreme scale.
Yu	Xiong	Case Western Reserve University	Cyberinfrastructure for bridge scour risk management.
Zhang	Yongjie (Jessica)	Carnegie Mellon University	A parallel computational framework of multiscale geometric modeling and mesh generation for cardiac biomechanics applications.

Appendix D

Collaboration Activities at the Workshop

As an optional activity, workshop attendees were invited to complete a *Collaboration Card* that was posted on a public posting board at the workshop venue. The purpose of this was to allow attendees to solicit new collaborators for joint research to facilitate networking beyond person-to-person interactions. A sample of the completed collaboration cards are summarized below:

Table 2. Sample of completed *Collaboration Cards*. Attendees listed both their own research interests and the areas in which they were seeking collaborators.

Researcher	Research Summary	Looking for Collaborators in:
Christian Ott	Blowing up massive stars and related topics in computational astrophysics and gravitational wave science.	Getting codes to scale, becoming architecture agnostic, getting better numerical methods.
Baskar Ganapathysubramanian	Modeling with multi-physics phenomena during fabrication of evaporation. Solving electrostatic equations.	Data mining, pre-conditioners, adaptive meshing.
Eric Wang	Using program analysis techniques to automatically detect programming errors and performance problems.	People who need automatic tools to debug code for coding and performance problems.
Gabriel Dos Reis	Programming languages and tools for dependable mathematical computation. Formal methods, symbolic and numeric computation.	Dependable systems, integrated software and hardware problems.
Xiong (Bill) Yu	Sensors, computer modeling of materials, fluid-structure interactions.	Materials modeling from solidification and microstructure sensor and data management.
Clint Whaley	Tuning computational kernels for extreme performance.	People needing extremely efficient computation.
Xiaolin (Andy) Li	Design adaptive scalable runtime management algorithms and software toolkit to support CSE applications.	Applications, load balancing and resource management algorithms, adaptive mesh refinement.
Kapil Khandelwal	Multiscale methods for topology optimization.	Optimization algorithms, distributed systems, sparse solvers.
Sophya Graschchuk	Quantum molecular dynamics, modeling chemical reactions including quantum effects in complex molecular environments, enzymes, materials, solutions.	Nonlocality in high (1000) dimension grids, meshes, interpolation, data clustering, random sparse matrix diagonalization.
Dario Pompili	In situ network data processing, supporting and running compute-intensive models in a distributed manner, extractive features online.	Real-time medical apps, ocean modeling and observation. Data center management and cooling optimization.
Diego Donzis	Massive simulations of turbulent flows. HPC, scalability, fundamental studies of turbulence and turbulent mixing.	Sharing data, “community” codes and allocations, exascale simulations.
Alison Marsden	Patient specific modeling for cardiovascular disease, optimization, UQ, multiscale modeling.	Open source management, image segmentation, meshing.

Table 2, cont'd

Suzanne Shontz	Development and research of parallel dynamic mesh generation algorithms. Theory and software for simulation-assisted medical interventions.	Mathematical modeling of medical devices and disease. Scientific visualization of device performance.
Xavier Trioche	Visualization at scale, structural analysis, data exploration.	Analysis and visualization in CSE problems. In situ visualization on HPC architecture.
Laurence Loewe	Build simulation model description language for systems biology, population genetics, ecology, and integrate with globally distributed computing to manage petascale data.	Data management, language development, simulation codes, solvers, visualization.
Ioan Raicu	Large scale distributed systems, including storage and computing.	Data-intensive applications, programming models.