National Science Foundation
Advisory Committee for Cyberinfrastructure (ACCI) Meeting
November 28th-29th 2011
Room 1235

Meeting Summary

November 28, 2011

Welcome, Introductions, Agenda Review, Logistics, Minutes

Dr. Larry Smarr called the meeting to order at 10:00 AM and welcomed the ACCI attendees. Brief introductions were made around the room and immediately following was an overview of the agenda. A list of attendees is in Appendix 1.

NSF Director Charge and Discussion

Dr. Cora Marrett, Deputy Director at NSF, spoke on behalf of NSF Director, Dr. Subra Suresh. It was reported that a budget was in place for fiscal year 2012. The importance of OCI working with other directorates across the foundation was discussed as well as the commitment to the CIF21 Initiative. The group was asked for advice and ideas related to a new NSF-wide strategy on high performance computing in an effort to maintain a broad and worldwide presence in a large spectrum of HBC activities and basic science. Dr. Marrett's report was followed by a brief question and answer period and discussion.

Committee of Visitors Report

The report from the September Committee of Visitors (COV) was presented by Dr. Taylor and Dr. King.

Dr. Mark Suskin presented the OCI Response to the COV report by highlighting several very important items.

Following the reports, the ACCI members had brief discussion and question/answer period.. It was mentionedthat in an effort to involve the community, the advisory committees should engage with other advisory committees and form joint working groups to help achieve that goal.

Overview of Cyberinfrastructure Activities

Alan Blatecky gave an overview of the current cyberinfrastructure activities. One of the major topics is directorate and office partnerships and how to identify projects and activities. In an effort to build national infrastructure, CIF21 is looking at co-development and deployment principles. On the topic of data, there are four issues that require attention: value

use, useful lifetime, distribution and data access. For software the focus was on developing software which is both robust and sustainable. After Mr. Blatecky spoke, there was a brief question and answer period.

CIF21 Programs and Planning in MPS, GEO and ENG

Dr. Sastry Pantula gave the CIF21 programs update on behalf of MPS. Many of the problems in the Directorate are very complex and require a grand community of players to solve. One example of this is MPS's new project entitled the Material Genome Initiative, whose goal is to produce advanced materials faster and at lower cost. Addressing these challenges requires a very comprehensive, sustainable and highly intelligent approach to CI. What is needed is an integrative, multidisciplinary grand-challenge program that enables communities to work together on problems that cannot be adequately addressed by individual groups. MPS has a webinar in December to share CIF21 opportunities with other communities, and they also started a CDS&E program. Additional items in MPS include:

- For the first time, an MPS category CDS&E for graduate research fellowships this year
- Working with EHR on training, especially for mathematical and science majors
- Participating in a workshop in December to connect internationally
- A new core program, joint with OCI, to support new ideas in cyberinfrastructure and CDS&E with hopes of expanding across MPS and other directorates at NSF
- A DCL and with several bullets related to amassing data, complex problems, visualization, data mining and cybersecurity
- An immediate need to develop core methods, technology and tools in computer, mathematical and computational statistical sciences in collaborations with other MPS sciences

With regard to HPC data and software strategies, it was felt that NSF needs to partner with other agencies for full-scale support as it relates to exascale capabilities, both for simulation and data analysis. NSF's current programs must be more sustainable and better support the scientific culture around the HPC centers. This will require long-term, sustainable support for funding.

On the campus level, the Campus Connections MRI-level facilities, midscale implementation and workforce development need to be supported. In Big Data, there is a need for data-intensive sciences creating a true paradigm shift that requires many new elements: data services, data curation, refined techniques, new analysis and knowledge extraction.

Dr. Cliff Jacobs spoke on behalf of Geosciences who felt it was important to continue and enhance cyberinfrastructure developments. In fact, GEO spends approximately $100 million dollars annually supporting cyberinfrastructure. One project is the Earth Cube activity. In an

effort to synergize activities along Earth Cube, it was important to create a dialogue with the community. With the support of OCI, a three and a half day charrette was held in early November that was attended by virtually 150 people. The event was productive and there are future plans to have another event in June and post a Dear Colleague Letter to further the ideas and concepts from the charrette.

Dr. Eduardo Misawa presented the CIF21 programs and activities for Engineering. The division stressed the importance of balancing research, education and workforce development because they are all very relevant to engineering. Some of the projects that Engineering has invested in are the NanoHub and NEES. Engineering is also considering having its own CDS&E core program to support research for single PI's and small teams.

After the presentations were complete, a discussion among the ACCI members followed which included items such as:

- How to devise a communication framework that allows people to communicate and share code and data of unpublished results; code data being available to regenerate and build on results;
- Requiring data access to published results in journals. (Due to the pressures to publish and "highlight grab," there's not much verification and validation on results, and it's hurting some sciences. Should some data be set apart and be used for verification and validation? And who will fund the activity?)
- The idea of a National Data Infrastructure;
- Large economic aspects of big bandwidth.

<u>CIF21 Programs and Planning in BIO, CISE and SBE</u>

On behalf of BIO, Dr. Peter McCartney spoke on the current CIF21 activities within the Directorate. The organization is driven to work closely with OCI. BIO would like to see growth in "cyberinfrastructure centers" in support of high-priority research. iPlant is one project currently funded by BIO which, to some extent , is serving in this kind of capacity. A Dear Colleague Letter was issued recently by BIO asking people to identify cyberinfrastructure challenges and to put together a conceptual plan describing what a center would do and what area of science it would advance.

Dr. Suzi Iacono gave an overview of CIF21 programs and planning for CISE. One of CISE's biggest challenges is not only collecting and storing mass amounts of data, but actually analyzing the information and making it useful. Some examples given were predicting earthquakes, tsunamis, tornados, bridge disasters, etc. CISE came up with two major areas of focus for the upcoming fiscal year. The first area is data analytics and the second is working closely with OCI to develop computational infrastructure. In addition, CISE had several other areas of interest including:

- Continuing to invest in DataNet
- A new Software Institutes Program to foster cross-disciplinary work
- The importance of developing new data mining techniques, machine learning techniques, new statistical techniques, new algorithms, new software and new visual tools

Dr. Myron Gutmann presented SBE's report on CIF21 initiatives. He explained that one of the primary needs in SBE is to have a cyberinfrastructure that will link data across varied sources by people, places or firms that will allow for data exploration and examination. One of the main activities that SBE is working with OCI on is improving the interoperability of SBE's large data surveys of populations, and the metadata from those surveys. Looking at ways to continue to develop these activities, a report was issued looking at the structure of what's happening in the sciences. The report concluded that it is important for communities to form up around the most important areas to build up infrastructure. After the SBE report, there was a short discussion among the meeting attendees.

CIF21 Programs and Planning in OPP, OISE and EHR

Dr. John Cherniavsky delivered EHR's report on CIF21 activities. In education, EHR continues to promote STEM to improve science and research. There are four connections that EHR sees with CIF21 and they are:

- New and effective tools for teaching and learning
- Big Data networks to allow for lots of collaboration
- Expanding capacity to undertake secondary analysis of large databases; statistical datasets
- Opportunities for collaboratively building a cyberinfrastructure workforce

OPP's Dr. Kelly Falkner updated the ACCI on its CIF21 activities. OPP is very excited about the CIF21 initiative and possible ways of incorporating cyberinfrastructure into its division. They have established an Arctic Observing Network. And in order to take it to the next level, there are some cyberinfrastructure challenges that have to be met. These challenges stem from international data linkage and accommodations. OPP is looking to network with different groups to figure out how to best process, handle and use data to enable science.

Dr. Scott Borg then further elaborated on different ways that cyberinfrastructure could improve OPP sciences. He also stated that potential for greater involvement in computer sciences and related issues included better tools for data mining and data compression.

The Office of International Science and Engineering's Executive Officer, Dr. David Stonner, spoke on behalf of OISE. He expressed that a major dilemma in OISE is not having the mechanisms in place to facilitate scientists who are already engaged in international collaborations, or who wish to engage in them. The introduction of SAVI (Science Across Virtual Institutes) in October, is a way in which NSF funded researchers can engage their

counterparts in other countries by getting a supplement to facilitate that engagement. A Global Venture Fund program was established recently and allows NSF grantees who wish to study abroad a supplement to their award for international engagement.

Update on Advanced Digital Services

The Advanced Digital Services update was given by Dr. Barry Schneider. He began by giving a history on the XSEDE program. The XSEDE solicitation consisted of two parts: high performance remote visualization and a data analysis. It provides coordination management, collaborative support and training, education and outreach. It consists of a technical audit service which is designed to provide information to people about what's going on in our digital environment. One of the difficulties is authentication mechanisms. And the division has adopted a mechanism for identity management and sharing, and allows its users to use their own local credentials. At this time, they are also offering consultation services to campuses as they develop their Cyberinfrastructure.

Discussion

The floor was opened up to discussion. At the concluding session of the ACCI meeting in June 2011, the Committee found that good progress had been made with the fifty-nine recommendations from the taskforce. It was suggested that the ACCI receive a report specifically on those recommendations. Dr. Suskin agreed that at the next meeting there would be a database of CIF21 activities across the foundation. The COV report had been emailed to everyone by the end of the meeting and the meeting was recessed.

November 29, 2011

Introductory Remarks

Dr. Larry Smarr welcomed the ACCI back for Day 2 and opened the floor to remarks. One observation made was the present and increasing interest of private companies in advisory committee meetings such as ACCI. However; there is usually little to no representation at most NSF meetings. It was recommended that more representation from private companies be present in ACCI meetings because it often brings significant and positive change.

Dr. Smarr proposed to the ACCI that the COV report be endorsed and all were in favor of the motion to approve the endorsement.

Dr. Suskin offered a short summary on some of the ideas behind CIF21, and that main goal is to support scientific research. There were several other factors that included sustainability and integration. It was also noted that OCI's $23million of CIF21's monies will be split between the four core areas: SI2, Data Net, XSEDE and Community Research Network.

OCI CIF21 Update

Dr. Rob Pennington gave the OCI CIF21 update on Data. OSTP is looking for input for two working groups that are looking at two particular issues: how one can publicly access scholarly information, and data generated to federally funded research. Also, maintaining and curating data is expensive.

Dr. Gabrielle Allen gave an update on software activities related to CIF21. Software is important because it connects together computation, experiment, theory and is pervasive throughout cyberinfrastructure. There are many challenges in developing and sustaining software in an academic environment

Dr. Kevin Thompson spoke about activities in Networking and Cybersecurity. Developments and activities that include expanding networking capacity in the IRNC program in FY 2012, including new ten gigabit connections and new transatlantic and transpacific connections. A new partnership with CISE has also been formed in cybersecurity as part of the Secure and Trustworthy Cyberspace program.

Ms. Mimi McClure spoke on behalf of Learning and Workforce Development on their current activities as they relate to CIF21. The main recommendation that came out of the report was the idea of Continuous Collaborative Computational Cloud, which is using interactive devices to bring people and provides access to data capabilities in the classroom and research lab. Another recommendation is to promote the cross-disciplinary transformative research in the understanding, learning and research mechanisms that make use of cyberinfrastructure for professional development. Other activities consist of:

Dr. Susan Winter spoke on Community Research Networks for CIF21. The emphasis is on improving the effectiveness of research communities which are now more diverse, larger and more geographically distributed than in the past. The focus is on understanding those communities and how they operate, to build and support them.

Blue Waters

Dr. Irene Qualters presented the committee with the Blue Waters update. The goal of Blue Waters is to sustain transformative science using petascale computation. An interim system is to be delivered in February 2012 for science using beginning in March. The floor was opened for discussion on Blue Waters. A concern was raised that no one had detected the difficulties IBM was having long before the announcement of the withdrawal of the IBM contract and also that there was insufficient information given to the community as to what was happening.

Models for Robust Software Sustainability

Dr. Craig Stewart gave a very brief discussion on the Models for Robust Software Sustainability. Along with his colleagues, a proposal was submitted to NSF, under the EAGER mechanism, for a case study on software sustainability; a restrictive scope of software. The goal

is to look at very few successful projects that have been around for a substantial period of time and determine what lessons can be taken from them. The group was asked for suggestions on any successful software developments, considered good sustainability models that should be looked into as part of the case study.

Software Selection Criteria and Issues

Dr. Nathan Bohlmann gave the presentation on software selection criteria and issues. The project is a quantitative study on the content of software CI and exploring approaches and methodologies that can be looked at to categorize and develop an inventory of what software investments have been made at NSF. Data Reproducibility

Dr. Victoria Stodden spoke on data reproducibility matters with the group. The focus of the project is on journal policy, particularly issues of reproducibility and knowing journal policies and journal publication requirements requiring code and data. The project ended up with 170 journals. So far 11 journals were using the term "reproducibility".

Announcements and Action Items:

- Dr. Craig Stewart announced that he was leaving the ACCI and that this was his last meeting
- Several items were requested from OCI for the next meeting:
    - An update on the evolution of the HPC program
    - A summary of how the ACCI task force recommendations are being implemented by programs at NSF.
    - Discussion of the strategic planning for engaging large facilities
- Dr. Craig Stewart announced that he was leaving the ACCI and that this was his last meeting
- It was announced that the next meeting in June 2012 will be a virtual meeting followed by an in-person meeting in late 2012.

With no further business, the meeting was adjourned at 11:45 AM.

Larry Smarr, Chair

3/12/12

Date

Appendix 1

ATTENDEES

Committee Members Present:

Dr. Larry Smarr—Committee Chair, ACCI; Director, CalIT2, University of California-San Diego
Dr. Shenda Baker—Professor of Chemistry, Harvey Mudd College
Dr. Thomas Diettrich—Director/Professor, Intelligent Systems Research; Oregon State University
Dr. John King—Vice Provost for Academic Information, University of Michigan
Dr. Michael Stubblefield—Vice Chancellor, Southern University and A&M College
Dr. Richard Loft—Director of Technology Development, National Center for Atmospheric Research
Dr. James Rice—Director, South Dakota EPSCoR Program; South Dakota State University
Dr. Victoria Stodden—Assistant Professor, Department of Statistics; Columbia University
Dr. J. Tinsley Oden—(via teleconference) Associate Vice President for Research; Director, Institute for Computational Engineering and Sciences, University of Texas-Austin
Dr. Todd Vision—Associate Professor, University of North Carolina at Chapel Hill
Dr. Craig Stewart—Associate Dean, Research Technologies, Indiana University
Mr. Alex Ramirez—CEOSE Liaison; Executive Director of Information Technology Initiatives, HACU National Headquarters
Dr. Elizabeth Lyon—(via teleconference) Director, UKOLN; University of Bath
Dr. David Keyes—(via teleconference) Foundation Professor of Applied Mathematics, Columbia University

NSF Staff and Other Meeting Attendees

Dr. Cora Marrett- Deputy Director, NSF
Dr. Jim Bottum- Vice Provost and CIO, Clemson University
Dr. James Hack- Oakridge National Laboratory
Mr. Alan Blatecky- Director, OCI
Dr. Nathan Bohlmann – Professor, Clemson University
Mr. Kevin Thompson- Program Officer, OCI
Ms. Mimi McClure- Associate Program Director, OCI
Dr. Peter McCartney- Program Director, Division of Biological Infrastructure
Dr. Cliff Jacobs- Expert, Directorate for Geosciences
Dr. David Stonner—Executive Officer, Office of International Science and Engineering

Dr. Kelly Falkner- Deputy Director, Office of Polar Programs
Dr. Barry Schneider- Program Director for TeraGrid/XD, OCI
Dr. John Towns – National Center for Supercomputing Applications,
Dr. Mark Suskin- Deputy Director, OCI
Dr. Valerie Taylor- Department of Computer Science and Engineering, Texas A&M

Appendix 2

PRESENTATIONS

COV Response (Suskin)

Overview of Cyberinfrastructure Activities (Blatecky)

CIF21 Directorate planning

- MPS
- GEO
- ENG
- BIO
- CISE
- SBE
- OPP
- OISE
- EHR

XSEDE Presentation

OCI CIF21 Updates

- Data
- Software
- Networking
- Learning and Workforce Development
- Community Research Networks

Blue Waters

Models for Robust Software Sustainability

Software Selection Criteria and Issues

Data Reproducibility