ACCI Meeting Notes

2 April 2014

Members present: Linda Petzold, Juan Meza, Colin Stultz, Kerstin Lehnert , Henry Neeman, Fred Choobineh, Lee Allison, David Bader, Jerry Ostriker, Victoria Stodden, Thom Dunning, David Yarron, Kenneth Bollen

Members participating via telecom: Peter Cummings, Deidre Evans, Sharon Glotzer

12:05 pm – Victoria Stodden called meeting to order. Reviewed minutes from last meeting- no revisions suggested. Linda moved to accepted minutes as is. Committee approved.

Irene Qualters (I.Q.) introduced and welcomed new members: Dr. David Bader and Dr. Thom Dunning (also an MPS AC member).

I.Q. went over agenda Highlights.

- Major upgrades to national and international infrastructure
- Updates
    - CIF21
    - Public access

I.Q. offered two science hors d'oeuvres described by ACI POs

- DAN KATZ: LIDAR – BICEP 2: current theory, the universe went through an expansion period, gravitational wave detection would demonstrate the accuracy of this
    - Two telescopes, B-waves
    - Recent announcement: polarization happening at 2-4 degrees in sky- the only theoretically-backed explanation is that this is a reflection of the big bang.
    - John Kovac, PI on BICEP 2
        - Supported variously, including at NSF through a CAREER co-sponsored by MPS/AST, MPS/PHY, CISE/ACI, GEO/PLR
        - Software was designed from this, but is reusable
        - Want to run more experiments, use software again, compare to results from other researchers, share data, etc.
        - Jeremiah Ostriker- yes, correct, results must be replicated. But this will be easy because software is now ready and right out of the box.
            - If this turns out to be correct, this is very important, defines cosmology completely.
- BOB CHADDUCK: Developing Sustainable Data Discovery & Interoperability Solutions- DataONE
    - Broad, trans-disciplinary interests, across disciplines.
    - Exemplar project across these areas.

- Sustainability of CI, understandings and activities advance because
- Federated model of coordinating nodes
- DataONE implemented across diverse institutions on a global basis. Scalability, extensibility.
- Engages broad communities- academia, federal sector
- Ex: DataONE contributed to OSTP's National initiative for safety data (hydrology data), as an important resource for public safety. Provides a portal to public to access NSF-funded data.

Dr. Farnam Jahanian, AD for CISE- NSF UPDATE

- Welcome, thanks on behalf of the ADs. NSF more than any other agency is community driven- review process, setting priorities, etc. Your work is an important component of that.
- Dr. Córdova will be sworn in this afternoon, committee will take a break to attend.
- CISE Mission (exploring the frontiers of computing; three objectives)
- Longer NSF budget presentations online for the taking.
- ACI Update
  - ACI not only advances frontiers but serves an integrative role across Foundation
  - ACI Division Director search committee instituted a broad external search. Just announced that Irene Qualters has been appointed.
- Two new Division Director Search Committees (AC members are welcome to send their input)
  - IIS
  - CNS
- CISE handled almost 8,000 props in FY2013
- CISE has also had a very integrative roll.
  - 61% of awardees are CISE disc
  - 25% from other sciences, humanities
  - Q (Jerry) has the number changed?
    - FJ: with ACI added, CISE disciplinary representation went from 65% to 61%.
- President's budget request – $7.255 B, approximately flat, CISE budget is flat.
  - However, between FY13 and FY14, CISE had a $36M increase
- CIF21 - $125M, $80M from CISE; emphasis on building communities
- Federal Big Data R&D Initiative that goes beyond NSF
- SaTC – CISE, SBE, and others
- BRAIN – growing program, broad ramifications, 3 agencies involved (and every NSF Dir)
  - $29 in BRAIN
  - Overall investment in neuroscience much larger
  - Advances in data intensive science will be critical
- Cyberlearning and future learning technologies

  - $37M request in FY14
  - Cross-disciplinary: SBE, learning science, education, CISE, and others
- NSF Research Traineeship (the new IGERT)
  - Goal same as IGERT
  - Primary research highlight of Data-enabled Science and engineering
- Many other NSF-wide initiatives that ACI/CISE contribute to.
- Advanced Cyberinfrastructure comprise of:
  - HPC
  - Networking, Cybersecurity
  - Data
  - People
- President's Opportunity, Growth and Security Initiative
  - Additional $552M.

12:48 AM: Additional members (on phone): Sharon Glotzer, Peter Cummings, Deidre Evans

12:48 PM Denise Caldwell, DD, MPS/PHY: The BRAIN Initiative

- A surprise that showed up in President's budget request
- It started out as brain mapping, integrated technologies
- NSF has, over quite a number of years has a significant investment in CogSci, neuroscience, BIO, ENG, CISE, SBE…
- What part of neuroscience is BRAIN?
  - Groups identified what BRAIN-related activities already being funded
  - Defined BRAIN at NSF as 5 thematic areas
    - Multiscale integration of the dynamic activity and structure of the brain
    - Neurotech and research infrastructure
    - Quantitative theory of modeling of brain function
    - Brain-inspired concepts and designs
    - BRAIN workforce development
- From molecules to behavior
- CISE, ACI communities will have significant input into the BRAIN Initiative
- NSF BRAIN Initiative website just went live.
- 12:59: conclusion of prepared remarks
  - Q: What exactly is BRAIN? What is NSF's unique contribution?
  - A: disease not a focus
  - Q: but NIH has programs for this, too
  - A: NSF has a unique role at the front-end. Things not tried or true, new measurement techniques, the exploratory.
  - Q: More bang for buck at NIH?

- Q: Helmholtz SC in EU will serve primary role of computing, data access for BRAIN.
- Farnam: NSF is a unique mission. We are pretty comfortable with the compute/analysis model
- Denise: fills gap for what NIH doesn't fund.
- Q: Is there GEO involvement?
- A: There is a Steering group at OD level, directorates express own interest.

1:06 pm Kevin Thompson – Upgrades to National and International Networking Infrastructure

- IRNC, HPNC
- Post NSFNET – NSF has been funding this for 20 years
- CC-NIE
  - Direct response to ACCI Task Force report from 2011
- IRNC
  - Graphic- not all connections shown. Other networks coordinated by other nations
  - Translight pacific wave: international links. – Mauna Kea telescopes, e.g.
    - 80 Gb between US and EU,
    - 2 10 Gb between US, Japan
- LSST: 3.2 Gigapixel camera, construction will start later this year
  - Tremendous CI planning on this project
  - Nighttime traffic in the 10s of terabytes, latency req's too
  - When this turns on at end of decade, this one data source will need 40 Gps
- Program not just about leased circuits and bits – science is end-to-end
- CC*IE- at least one rec from Taskforce report on campus bridging
- Transition of research prototypes in SDN, etc.
- CC*IE
  - 84 awards over 2 years
- No Questions

Mark Suskin, Tom Russell – CLG

- CIF21 a key profile- $124 M
- CLG representation from all directorates
- CIF21 : 3 foci
  - Community-building
  - Data-intensive science
  - Computational models, methods, algorithm tools
- Table of CIF21 activities
  - CIF21 a very cross-disciplinary activity
  - Q: Is GEO on Big data
  - Q: What does participation mean? Financial contribution?

  - A: Not necessarily up front, could be interest if a relevant proposal comes in for funding, could be intellectual contribution to the solicitation. This is an ongoing process – even if CIF21 does not persist as a named initiative, CI coordination is ongoing

- Tom Russell
  - NSF-Wide CIF21 Data Vision, 5 pieces that align with Farnam's slide (his BigData graphic)
  - Internal WG is trying to catalogue the principle data-related activities taking place in foundation across all directorates, map onto data vision – see what the interests of directorates, and visions are.
  - Document will be updated as an official 14- (note: as of October 2014 this is still a work in progress)
  - Catalogue will be used to identify overlap, common interests, help to coordinate, better characterize data-related initiatives for the community
  - Also working on directorate-specific documents
  - Q: What are "community-based difficulties?"
  - A (Russell): for example, reward systems and data sharing/best practices

Discussion

- Jerry Ostriker: Tradition in astronomy is for researchers to collect data and put it in a drawer. SDSS turned that on its head. Can discuss offline how that came about, but this is like any game: Those with the resources set the rules.
- Kerstin Lehnert: Shared infrastructure yields homogeneous data. Specialized equipment makes it hard to encompass that homogeneous, data sharing attitude
- Jerry Ostriker: It looks this way because of the changes. Often there is a unique researcher that owns his/her data, but it just depends.
- Collin Stultz: some data can't be moved because it is too large. Problems with creating a common infrastructure challenged by inability to move or restrictions on moving data. Fear v. money – two main tools? Combining data from different resources is the way to think about it.
- Algorithms, software codes – maintenance is an unresolved issue. Any intentions for maintaining software?
  - Dan Katz – That's what the SI2 program does. All parts of NSF participate. All communities see this as a common problem. People see use in a piece of code, but much effort/funding is required to update codes as infrastructure changes. 3-5yr funding for this, asking researchers to think of plans from the future.
- Most software has lifetime of 20-30 years
  - DK – Yes, there is an option for them to come back and get more funding
- Thom Dunning: software institutes also a way to build communities
  - DK – There are three types of awards, development, institutes (community-forming), …
- IQ we are talking about sustainability of software investment and data infrastructure
  - Does ACCI have examples to call out for strengths and weaknesses

  - Kerstin Lehnert: Australian National Data Service
  - Jerry Ostriker: Requiring data access v. effective data use, must be put in readable form
  - Collin Stultz: NIH has a requirement, doesn't work well at all
  - Thom Dunning: Chemistry has a history of making codes available. But very little new computer science used. SI2 a way to break through that bottleneck, but what will happen in 10 years as the computing technology continues to evolve
- David Bader: Looking towards the future is more than collecting data sets – access is interesting, but more and more we must look at the actual analytic use of these datasets
- Kerstin Lehnert: Formats are significant in utility
- Tom Russell: Are there communities whose CI needs are not currently well-represented?
- Lee Allison: EC governance: concerned communities not ready – starting at ground zero, no common standards, no interoperability. Synthesis of EC workshops should be ready in next 2 weeks, will be put online
- Kerstin: it is clear that the community building needs incentives for people to participate. Resources are distributed – participation of individual researchers hinges on "is this the correct format" – Small resources, spread out widely could have a big impact.
- Tom Russell: How should we do this?
- EC POs have been explicit that they will not support databases right now. Very unclear where they should go
- Eva Zanzerkia (GEO/OAD): One issue is communicating this properly to communities – our concern is that we don't end up with piecemeal databases that are not interoperable.

Overview of MPS Facilities and Cyberinfrastructure Working Group

- Wayne van Citters, MPS/OAD
  - Provide context, including description of the resources that MPS facilities provides
  - CI elements of facilities portfolio
  - Challenges to ACI collaborations
  - MPS community is diverse and has diverse needs for facilities
  - Math and chemistry have small facilities needs relative to AST, PHY, and Materials
  - Facility vs. individual investigators
  - Described ART, PHY, DMR facilities
  - LHC is largest existing collaboration/database/cyber challenge that we have faced; it was very successful.
  - ALMA, LIGO
  - Two different budget lines for infrastructure
    - R&RA- maintenance, networking, software
      - ALMA
      - IceCube
    - MREFC

        - Not to exceed costs
    - Funds cannot be mixed after the fact
  - Challenges
    - Not invented here
    - Natural desire for local control
    - Uncertainty about continuity of service
    - Different cultures
    - Mutual lack of awareness (about what ACI can provide)
    - Flat budgets- priorities given to staff and instrumentation
- Bogdan Mihaila, MPS/PHY
  - On MPS CIF21 MPS group
  - WG was formed for the directorate
    - Came up with a roadmap, includes data, software, hardware, compute, workforce development
    - A multitude of activities address these
    - CDS&E
      - CMMI, CBET overlap
      - A forum for overlap between programs in C&D
      - More than 400 proposals, 30% were submitted to WG for possible co-review. 5% of grand total were cofounded – not certain whether these would have been funded otherwise.
  - Our collaboration with ACI is absolutely critical
    - Domain scientists focused on solving problems – we are not always doing the best computations, often use the quick and dirty approach that gives us results
    - However, other programs, like molecular mechanics, at some point become larger than the PI
    - At some point there must be a handshake to ensure we are not reinventing the wheel or wasting time, and make sure that the codes we write are efficient and consistent with the ever-changing architecture.
    - Very much interested in data, hope to be
- Jerry Ostriker: The pyramid is inverted – now more money is going to data/analysis than for the telescope; this is the future for many projects.
- I.Q.: is there interaction between facilities and the CLG WG on these? There could be some bridging between the culture and the language
- Thom Dunning: Collaboration between MPS and ACI is important; MPS wants to be able to do a lot more science with the money that they have.
- Juan Meza: On this mutual lack of awareness – this sounds like a serious concern that should be easily addressable. Are there plans for addressing this? Is there something the ACCI can do?

ACCI Meeting Notes: 2-3 April 2014

Irene Qualters

- 2013 was big year for resource deployment at high end, e.g. Blue Waters, Stampede, and NCAR (for GEO)
- Resources are increasingly collaborative and diverse, i.e. built for different purposes such as for visualization, big data, and with large memory
- Some machines are getting shut down as well, but this is deliberate
- Number of institutions using resource pool has doubled although the pool is smaller overall over 10 year period
- Number of projects running on resources has tripled (200-1200) in 10 years
- Number of active users has increased 5-fold
- Number of requests per allocation has increased 2-5X versus what is made
- Total of core hours has increased and eventually plateaued; when normalized this is normal growth
- All scientific disciplines have increased use of resources broadly-speaking
- Increase in use of resources is broadly distributed
- Use of resources is well-characterized
    - Additional analysis is beginning
    - Modest increases in parallelism
    - Emerging scientific and workflow trends
    - Role of industry?
    - Full costs are highly competitive

Jon Eisenberg, NAS

- Study is underway
- Wants ACCI input
- Walked through slides: charge to committee, planned study products, committee selection, draft questions for interim report… science needs and opportunities/advanced computing capabilities facilities and requirements/challenges and suggestions, inputs include web conference and written comments

Robert Leland, OSTP

- National strategy for HPC
- HPC is a strategic advantage for the country, we have a 5-7 year lead
- This is in national security, economic competitiveness, and scientific discovery
- What does the government need to do to sustain this?
- Strategic drivers:
    - International competition
    - Rise of big data – how do we adapt our thinking to deal with this
    - Erosion of Moore's law

        - End of stability we've enjoyed through massively parallel computing period of the last 20-25 years; change in the programming paradigm to some new approach
    - What ought the government do to help the country move forward given these drivers?
        - Whole of government approach
        - Must be strategic
        - Broad view of what computing means, i.e., this includes advanced networking infrastructure
        - This is an initiative, we want some effort above the baseline
    - How can we link the agencies and how can we lift the effort
    - Primary needs are around simulation, cryptanalysis, and analytics
    - Industry echoes this but with more emphasis on analytics

Tom Jordan, USC

- GEO is becoming a system science
- Must forecast appropriately, which requires large ensemble computations
- Lots of data into big models means need for HPC to forecast
- Walks through slides: CA situation, workflow for how it happens, cybershake simulation, cybershake workflow, simulation needs, GEO drives vertical integration of CI layers, SCEC needs reflect needs of GEO broadly, validation of simulations shows improvement in models, procedures for gathering HPC resources, recommendations
- Last recommendation – need formal collaboration between DOE and NSF
- Irene Qualters: SCEC runs on different types of architectures, how do you mediate the differences in architecture. How big an issue is it?
    - There's not Moore's law for software. SCEC adapts and develops new software to use different systems more effectively
- Collin Stultz: What about specialized hardware that is not generalizable? Can you take advantage of a common bottleneck to build specialized machines?
    - SCEC calculations are varied and system focused. It's necessary in this case to have large scale facilities
- Jerry Ostriker: Open ended questions about how NSF CI can do a better job. Can other communities make suggestions for how to do a better job? This is aimed at Eisenberg. For example, software communities focused on extreme scale solutions.
- Thom Dunning: how do communities see a need to allocate investments? This is also for Eisenberg.
- Farnam Jahanian: Priorities. Also, some programs try to deal with software issues.
- Thom Dunning: Use of parallelism has not increased significantly (according to Irene's data) and this is due to limitations of software.
- Rob Leland: We're heading into an area of increasing hardware diversity, highlighting the importance of figuring out what to do about software.

- Juan Meza: is there an awareness at OSTP, NSF that you need fundamental research 5-10 years out before they can produce algorithms that lead to software for new HPC
    - Rob Leland: OSTP is aware of this. One aspect of initiative is to look over the Moore's law horizon. Expect the transition to this to take over a decade.
- Farnam Jahanian: Investments at IARPA, DARPA, elsewhere are being made in the area of algorithms. But what are we going to do in the next 5-10 years. OSTP report must really look at spectrum of time frames. Farnam gives strong endorsement to two activities, OSTP and NAS.
- Irene Qualters: consider these questions while talking to the ADs for their disciplinary perspectives.

AD meeting

Fleming Crim, MPS; John Wingfield, BIO; Pramod Karghonekar, ENG; Joan Ferrini-Mundy, EHR; Farnam Jahanian, CISE; Joanne Tornow, SBE (Acting)

- Thom Dunning: what are your needs, what are your challenges
    - JT: Big data and analytics, admin, survey, observational, neuro (e,g., fMRI), quantitative approaches to social science, melding of qualitative and quantitative data
        - Partnering with ACI on BCC-SBE/EHR, DIBBS; partnering will be the likely approach going forward
        - One challenge for SBE is that there is not just one community but many different types of data and ways of collecting it
    - FC: many disciplines, many types of support provided, all turn on CI broadly defined
        - Computational techniques, algorithms… these are all CI to MPS communities
        - BICEP2 is a good example of how the foundation works collaboratively
        - LSST is an enormous dataset with a telescope attached to it: it is a big data problem, networking problem, access problem
        - LIGO is going to have similar data issues as LSST
    - Thom Dunning: What plans do you have for engaging ACI to meet these CI needs?
    - JW: Exploding amounts and diversity of data in BIO… we went from one human genome 10 years ago to thousands of genomes now
        - Digitizing natural history collections (dark data) along with tool development by iDigBio
        - iPlant facilitates data enabled bioscience that is entirely new, also serves somewhat as a connection to HPC for the BIO community
        - Workforce preparation issues are a huge challenge for the BIO community, especially for data science
        - Workforce preparation also fails to get students to program at a basic level; more and more of this is happening at the graduate level but not as much at the undergraduate level

- PK – Engineering research has developed alongside computing research. Nanohub is one great example of a successful ENG/ACI partnership. Also, NEES. ERCs have sizeable CI needs. RIBS solicitation is thinking about infusion of CI into systems and also on interdependent systems.
    - Materials Genome Initiative is also an example that will have strong CI needs and is interdisciplinary and interagency
    - ENG has a close working with ACI and expects this to proceed into the future
    - IQ: Are there specific workforce needs in ENG that must be addressed. There may be some best practices to be gathered from a few examples that can be shared broadly.
    - LP: It would be great to demonstrate the value of programming to students, who are motivated when they understand the need for and value in an area.
- JF – Role of CI in teaching and learning; how can we learn more about and better understand issues of teaching and learning in the context of CI
    - NSF can experiment/pilot with different ways to provide skills and knowledge to students that might use CI skills
    - NSF can stimulate these experiments and learn a lot about what different approaches at the UG and grad level are working to improve learning about CI
    - Improving Undergraduate STEM Education program and the new NRT program both have provisions for CI learning, including
    - Want to support teaching and learning both in and out of school – innovation in learning technologies, tools for learning, resources for learning, e.g., online games, museums, after school resources.
    - What resources are made available to teachers? How can teachers use the enormous number of resources online, judge them, and work with them in efficient and effective ways.
    - Can we use CI to understand better teaching and learning? Not as interested in supporting development of MOOCs but rather those that are testing teaching and learning in new ways. There are EHR programs that will fund activities of this sort.
    - Use of video data to study teacher-learner interactions. What to do with the proliferation of data sets that might get us more quickly to answers about teaching and learning. There are many data sets that exist… what new questions can we ask and how can we use these resources?
- Colin Stultz: What is meant by CI? Seems that different directorates mean different things. Size, diversity, linking of data means different things in different disciplinary contexts.
- JW: We do try to work collaboratively with other fields, e.g., Earthcube.

- IQ: Common theme – Learning and workforce development. Another common theme that comes up is software sustainability. Foundational level – big data activity. Also, what are the basic algorithmic approaches to attacking science problems? Software institutes also cut across disciplines and touch on common themes.
- Ken Bollen: Cross cutting desire to open up data to as many people as possible. But we also want to restrict certain data. How can you do both simultaneously?
  - Open data and public access (both OSTP memos)
  - CISE has cross foundational cybersecurity program, and CISE/SBE has CFP on what privacy means and related topics
  - Stodden co-edited book on privacy and data (http://www.dataprivacybook.org/): what's legal to collect? What are the controlling laws here? What are our goals as researchers? Call in book to be more innovative on how to permit reproducibility and reuse of data even when there are barriers to sharing (legitimate or otherwise). The goal is to get to a point where you can obey the law while sharing data for validation or verification of research.
- JW: DURC (e.g., H5N1 research)… should this be shared broadly? Synthetic biology makes this a real concern for misuse of the technology. Debate on how this is going to be managed.
- FC: Must share data that will have a potentially long life. Archiving this is a real problem. How can you keep the archived data linked to an up-to-date technology? What about the hardware considerations? Who is going to curate the data?
- Kerstin Lehnert: Many communities have NSF funded facilities for data. It's an ongoing question about this… do you fund these out of research programs, out of infrastructure, how do you do this?
  - FC: if you create something, you have to operate it or curate it. This needs more thought.
  - Farnam Jahanian: This is an issue that we have no broad solution for. There is an evolution of CI that we are seeing, from hardware to computation to data.
- IQ: Data stewardship – need retention policies that are policy driven. Whose data do you keep, and what do you let go of?

Discussion in preparation for meeting with NSF Director and Deputy Director

- Explain the role of ACCI
- Do introductions to members
- Anything committee would like to share with the Director
- Hear from the Director anything that she would like to share
- Dr. Córdova was on the NSB strategy subcommittee so she may be able to talk about that
- Juan Meza: what is her vision for NSF – this may be a good question for the Oct/Nov 2014 meeting

- Farnam Jahanian (FJ): What are your perspectives on NSF and CI to Dr. Cordova
- Colin Stultz: Should we talk about obstacles or why CI is important generally?
- Juan Meza: Plant a seed on the importance of CI
- IQ: Plant a seed on where CI matters most for NSF
- Farnam Jahanian: See trends and themes for CI, e.g. data
- Victoria Stodden: There are traditional ideas about what CI means, how can we talk about CI as something new
- David Bader: Interesting to hear her perspective on how US can interact internationally, both competitively and cooperatively. FJ: this may be a question for the future.
- Fred Choobineh: Why is it important to continue investing in this area? Why is ACI important? Then… what are the challenges in this area?
- David Yaron: Foundation-wide investment in CI is enormous. IQ: half of what ACI spends goes to other Directorates. FJ: this investment is growing as we invest in the CI component of facilities.
- Jerry Ostriker: Total foundational CI investment is much greater than ACI investment in CI for the foundation.

April 3, 2014

Victoria Stodden: Report of StatsNSF Subcommittee to the MPSAC

10:10 AM JOINT MPS-ACCI Meeting

Farnam Jahanian welcomed the MPS AC in joining the ACCI

- This is an experiment
- For the next few hours, think broadly, think big, use this as a brainstorming session. Everyone recognizes how CI is becoming increasingly important to science.
- Three broad categories that we invest in. Research & enabling technologies. Transformative applications. New algorithms & models. Various organizations invest in provision & deployment of advanced CI.
- When we talk about CI, we are also talking about people, educating the next-gen of computing-savvy scientists and engineers

Fleming Crim

- Very glad we are having these discussions.
- Important to have a discussion about needs, strategies, tactics, communication
- Asked MPS folks to prepare a list of interactions with ACI. A very long list ranging from cross-cutting initiative to individual investigator grants.
- To us CI is not just hardware In fact it is mostly not hardware- algorithms, techniques, resources for computing important – much of what MPS does depends on data and computing.

- Hoping not only that MPS consumes ACI's resources, but also that it contributes to the mission.
- We live and die by this stuff, and want this to be a fruitful interaction.

Before going into the breakout sessions, Irene thanked those who contributed to the topics

- CDS&E pipeline – Thom Dunning
- Research challenges – Juan de Pablo
- Coordination and leverage between MPS and ACI facilities and CI programs – Victoria Stodden

Groups departed to breakouts

12:30 PM Breakout Group report back.

Thom Dunning reported on the Pipeline working group

- We heard a lot about algorithms/codes, etc.
- Development of model, math & algorithms, computing system itself. Must support the full spectrum to advance science
- Group feels that we need to embed knowledge of computation/data, etc. in science curriculum. There are classes, but the current ones often don't give the scientists what they need.
- Teaching a way of thinking is important
- Content and curriculum – NSF could support development of new curricula/courses for all modern scientists and engineers (not just the computational ones!) – to redefine what it means
- K-12 curriculum has significant needs.
- Partner with Scientific Societies – for example, ACS approves degrees. The prestige and weight of the scientific societies (and their certifications) will have a bigger impact than NSF alone.
- Broader impacts in proposals? These could be used as small, incremental steps.
- All of this will take time
- NSF could sponsor the development of short courses and summer schools. In Thom's experience, student enrollment is high, meaning there is a need!
- Facilities professionals have a significant body of knowledge that could be tapped. This group should be better engaged, whether in short courses or through mentorship.
- What is the body of knowledge that a computational scientist needs to know?
- Benchmark educational activities in other countries – how are they educating their students
    - Japan established a research center and a graduate school while developing the Kei computer
    - German research school in computational science and engineering.
- Fleming Crim: a number of these ideas should engage EHR.
- What can we do on the training front?
    - NSF can fund efforts to establish models
    - We do not have resources to fund the activity everywhere.

ACCI Meeting Notes: 2-3 April 2014

Juan de Pablo reported on research challenges

- Surprised about how similar the opinions of the MPS AC and ACCI members were.
- Everyone has his/her own code, software.
- In several disciplines (geology, climate) there are common codes written in this country.
- In Materials, most of the code is written in EU, we do not have them here.
    - Want to ensure we write, curate codes here.
- Where to invest resources for software development? Do we have mechanisms for curation?
- Lack of awareness of what is going on in multiple divisions/directorates, disciplines, and with other agencies. Better communication needed.
- Education a big need. MOOCs can enable students to rapidly come up to speed.
- Hardware:
    - Committee members do not want to see the science chasing the hardware, but rather, the hardware should address science's needs.
    - Though we do develop algorithms/new ideas when the computers are there.
    - Computational legal systems?
    - National center for software experts?
- No one wants to write/review extra proposals, but a supplement would be ok.
- Form a small subcommittee to raise awareness, provide ideas.

Victoria Stodden reported on coordination/leverage of NSF CI and facilities

- Tasked to look at overlap between programs.
- Data that came from MPS – no resources to make this persistent.
- Create an NSF group to support data persistence
    - Begin by funding pilots
- Software persistence
- Data, software:
    - Versioning.
    - Interoperability between models, data sets, focus towards enabling new projects and new science.
- Can we learn anything by looking at pilots and investments with competition vs cooperation – which is a better model?
    - This should be across agencies.
- Need greater coordination around applications for compute time.
- Identify commonalities in existing MPS data management infrastructure.
    - How is software managed?
    - What testing is done?
    - Identity management?
    - Commonalities? If not, is this due to differences in the research itself?
    - Smaller scale projects?

- Simulations and data output:
    - Different types.
    - Archive software settings to replicate simulation data vs storage of output data?
    - Are older data even useful?
    - Is there a way to preserve the environment?
    - Software citation? Repository? Reproducibility?
- Jerry Ostriker: Most time spent on the idea that NSF through MPS might take on some of the job of data storage. NSF does this for some individual programs (OPP). NSF might
- Kerstin Lehnert: EC workshop brought together facilities in earth sciences, formed a council to enhance collaboration within facilities – many developments here may be used to inform other communities
- Phil Buxbaum: larger issue in other communities
    - Synchrotron data/storage. Before a single collaboration owned whole process – now it is becoming an issue
    - Data access is really a science problem – important to ensure that there is science taken into consideration
    - MPS will discuss this

Victoria Stodden adjourned the meeting *sine die*

Irene thanked the ACCI

Farnam thanked the MPS AC

Fleming noted that the conversation needs to continue, even beyond this group, back in the communities.