<div align="center">

Minutes
Advanced Scientific Computing Advisory Committee Meeting
June 23, 2011
Teleconference

</div>

ASCAC members participating via telephone:

Dr. Marsha J. Berger   Dr. Thomas A. Manteuffel
Dr. Jacqueline H. Chen   Dr. John Negele
Dr. Roscoe C. Giles, Chair   Dr. Linda R. Petzold
Dr. Jack J. Dongarra   Dr. Vivek Sarkar
Dr. Susan L. Graham   Dr. William M. Tang
Dr. James J. Hack   Dr. Victoria White
Dr. Anthony J. G. Hey

Other Participants:

Barbara Helland
Melea Baker
Christine Chalk
Daniel Hitchcock
Lucy Nowell

At approximately 2 pm EDT, Dr. Roscoe Giles called the meeting to order and announced that call was being recorded for minute's purposes. A roll call of members was taken and it was confirmed that a quorum of members were present.

Dr. Giles opened the meeting by thank everyone for participating. He noted that the purpose of the meeting was to discuss/review the data policy report that Jim Hack has been working on - for ASCAC to contribute to that report, adopt it in some form with edits and suggestions and vote to accept or reject the report. It was also noted that the committee was working toward a July 1 deadline to reply to the Office of Science on this charge. Dr. Giles then thanked Dr. Hack and his committee for all the hard work on the report – especially given the very short time to deal with a broad set of questions – to ensure the report is clear and effective. Jim Hack was then asked to briefly summarize the work of the committee and the draft report.

Dr. Hack noted that his committee tried to quickly aggregate input from the members and other sources. They found that a large part of what is currently done with regard to data policy is "ad hoc," in that it is not governed by policy, and the policies that are out there are somewhat dated.

Dr. Hack than outlined major findings and noted the need for something more thoughtful in terms of time spent to properly review this from a scope and depth point of view (how to maintain flexibility to manage the data and publication enterprise using available technology which is always changing; to encourage use of best practice to move fields forward; and the impact on workforce development – since the availability of data helps young investigators move

<div align="right">1</div>

state of the art research. Dr. Hack summarized the four finding: the first has to do with how to discuss the results of scientific research and publications as the primary mechanism for disseminating this information; the second has to do with scientific user facilities, particularly ASCR facilities and to what extent the data is the responsibility of the investigator; the third finding has to do with software and it's dissemination; the forth addresses an issue raised at a previous ASCAC meeting regarding community-specific issues – particularly communities that have strong support from the ASCR office.

Dr. Hack also noted that specific disciplines and certain research communities are ahead of themselves. "There are a lot of caveats here; we're just touching the surface of this. Other agencies have invested the efforts of dozens of people to thoroughly understand how the data in all forms is disseminated to the public. The Office of Science needs to develop a sustainable strategy. For a center director: we can't afford to store everything for all time; it's not feasible. Balance is needed between knowledge infrastructure and the investments/creation of that knowledge."

As a summary of this discussion, Dr. Giles led a discuss of the proposed report outline against the specific details of the Department's charge letter. He noted that the structure and areas were thoughtful and nicely arranged and reflect what is done today. Dr. Giles also commented that the Department of Energy Office of Science (SC) is trying to gather data to inform what the deeper study and policy should be in response to this mandate from the America COMPETES Reauthorization Act. "We're not being asked to figure this out ourselves, but to provide information to get the process stated. I don't think we need to go into this deeper except to say this is where we, and our communities, are with regard to data policies and dissemination practices.

Dr. Graham stated: "I agree with that, but I suggest that maybe we have an additional finding that this is an issue that transcends ASCR, even DOE: that a lot of groups are studying this, that it would be beneficial to have one coordinated effort rather than lots of individual efforts. I would like to see this at the NITRD level, at least for scientific simulations. Data is the hard part and is beyond NITRD."

Dr. Chen stated: "Data has been growing at huge rates. Most of the efforts in the Office of Science have been grassroots efforts, unfunded, in that we are funded to create the data, but there is no support for sharing and curating the data. Centers where data are generated are not responsible for retaining the data. The potential benefits to the overall community are lost unless individuals take it upon themselves to do this. It takes substantial resources to set up data sharing and dissemination and there needs to be long term support for maintaining that.

Marsha: "Which data is good for long term savings?

Jackie: "It's community-dependent. One size fits all is not going to work. Each discipline has its own.

Marsha: "[How scientists handle this issue is] evolving rapidly; we wouldn't want to prematurely lock into something. For the last paragraph: maybe our recommendation is that we

should monitor this and see what evolves before saying something. We don't know enough. A second point: in the findings what's not mentioned is who pays for it all. This is a big issue, maybe it didn't come out enough?

Jackie: "It is in the summary, but we can amplify that point.

Marsha: "It's in your example."

Vicki: "On the issue for… the interagency issue: the charter for that has ended. A new charter is due any day now. There are two interagency groups working on this issue.

Vivek: "The current practices are *ad hoc*. That's not necessarily bad. How we store data has changed dramatically over last 20 years? Leaving that up to specific disciplines on how to store and curate the information is good. We should make sure that "*ad hoc*" is viewed positively and encourage people to save data rather than waiting for a policy.

Jim: "I didn't use [the term] *ad hoc* in the report because of the negative context. It's not equal to a bad approach; it's one in which the community does not have resources to do this in a proper way. It's currently not the best use of software and hardware technologies. This becomes a "roll your own" type of approach. We need to help clients start to define how to support infrastructure needs in these communities.

Vivek: "We should identify communities of importance to ASCR. I'm also curious the tipping point of the science [regarding size of data set]. Too small, and it's a waste of resources. We need to focus the community where the investment really pays off.

[John and Vicki then announced that they had been on the call but had not announced themselves initially.]

Roscoe: "We need to turn the focus onto public access. Who is the audience is for this? What kind of question of broader public access should we consider? It sounds to me like what we're saying is: historically those communities have been most interested in being able to disseminate/ receive results of publicly funded research in ASCR, and they've worked to develop means of dissemination and rules that support those communities. We're looking at a different level, to enhance efforts within those communities but also the public dissemination and what's being shared across communities. [The request is to] codify something at the overarching level but not too early for the broader policies.

Vicki: "What the Brinkman letter seems to ask for is what Jim did. We're drifting into a discussion of recommendations that I started, but now I wonder if we should be doing that or not.

Roscoe: "Recommendations get turned into unfunded mandates.

Vicki: "…More in warning kind of recommendations. He did actually say that… I'm worried about a top down thing being imposed, becoming an unfunded mandate.

Roscoe: "What we're saying is existing dissemination are being developed within the individual disciplines. On one hand, it's good to encourage that and make it systematic and funded and talk about common problems, but it's premature to impose a one-size-fits-all mandate.

Vicki: "This is important."

Tom: "A comment on discussing data in our field. It's the software. This is a sticky problem. It's an issue that we shouldn't comment prematurely on.

Vicki: "Is that how ASCR-funded researchers fund their software? Does everyone who gets an award have to make their software [publicly] available? What about the issue of research advantage? I've done some reading, looked at workshops, etc. That's a very big issue.

Barb: "They have to bring their own funding; we provide our capabilities and expertise."

(Unidentified man): "This is not an unusual situation. Agencies will make tremendous investments in making observational measurements. The analysis is not part of the big picture plan. This is the generation of intellectual property, be it data or software, as Barb's suggesting. The real emphasis is on the front end, but the back end is ignored.

Lucy cited a current research policy regarding intellectual property. "We do have a requirement on ASCR funded property." She also made a point regarding copyright assertion.

Jim: "That's in the report."

Marsha: "What about, regarding an algorithm: you write your own...would that fall under that kind of policy?

Lucy: "The specific policy [is] with software developed with funding with ASCR.

Barb asked a question about maintenance and user support.

Lucy: "The policy is on our website.

Roscoe: "I put that in our email. The question on maintenance and user support is important: the issue of making software available.

(Multiple concurs)

Jim: "The software, maintaining it, helping people figure out how to use it: we don't have resources for that.

Marsha: "...making it available for others is not part of our funding.

Roscoe: "I know we say something about that in the findings; something on software could be called out.

(Unidentified man): "There is nothing at the current time. For SciDAC, fine, but in general, for algorithms that are developed, there is no current mechanism to support it or bring it to open source.

Barb: "If facilities use our software they will maintain it. For software developed in the research program, they have to maintain it.

Roscoe: [inaudible]

Vicki: "The issue is ASCR saying you must have an expectation of funding [for such a policy to be] usable. …Open source license. I don't know if that's what ASCR intended or if meant [make it] open source if the software is usable or potentially usable.

Tom: "I think you said it very well. Neither obligation nor funding to make [software] open source and making it be usable.

Vivek: "There are a number of open source licenses. It's about freedom for commercialization. Eclipse is an example of patent protection but that gives freedom for commercialization.

Lucy: "There are some exceptions for open source requirements from ASCR, but they have to be negotiated at the time rather than after the fact.

Susan: "That gets back to the issue someone raised earlier on data. Not all software produced is useful enough to warrant additional investment or to maintain over time. There isn't a selection mechanism that we understand, either.

Linda: "I'm thinking of software that is not of production quality. Here's an analogy: if I had some data and there was a problem with a chip or reagent, am I under the letter of law supposed to post that data? If I have suspicions over [the data]: over the interest of everybody, I don't post it.

Roscoe: "There must be some threshold in what's adopted by libraries. Those who are choosing to adopt it... I'm not sure there is overall guidance on developing software.

Vicki: "I think [one must consider if it's] usable by others. Your own code may not be at that state, but it's not incorrect. It doesn't mean the data that's produced is wrong.

Linda: "That's why I called it 'deficient:' [meaning,] it's not dissemination-ready, but it's not numerically incorrect.

Vicki: "One of the purposes of making it open source is reproducibility. You can do that without expecting software to be used for additional purposes?

Linda: "Who would be using it? Even that level…that I use on my machine… [it] is not always reproducible.

Roscoe: "In science that's the way it is. It's not always the way you want it to be. I build a custom counter; in the paper, I should provide information so others can reproduce that.

Tom: "Part of 'state of art' is there is this gap. If they want it to be fixed, they need to give funds.

John Negele: "I've been cut off twice. Christine: could somebody check and see what's going on?"

Christine: "I'll check."
Roscoe: "Sorry, John."

Roscoe: "If we can identify it, call it a gap, then "gap." Not fully published research results. But with software, I think we need to contain it right. It almost never is usable by another person, fully, without being maintained. I think my WordStar software worked well if only I could find a CD and machine to run it on.

Jackie: "A question for Jim – did you attempt to cross-link with the findings for other SC efforts? Have other offices identified large amounts of software generated from their work?

Jim: "A couple weeks ago, we were very interested where others stood on this. Roscoe got one of the subcommittees to send an executive summary on how they were tackling this. It had a flavor of things much more specific to that discipline. [The area was] nuclear security. The answer is no. For me, it was more curiosity as to how others were tackling this. With this problem, you can dig yourself into a very deep rabbit hole. We only wanted to try to respond to a broad guidance, to answer, "What are we doing now; what policies are in place?"

Roscoe: "In the very beginning, I sent a note to other chairs asking how they are handling it. Some appointed subcommittees; others tried handling as it as a committee as a whole. People were going in the direction of producing their findings more than recommendations. We must be cautious to not lead to unfunded mandates. We wanted to provide information requested by the office but be careful not to end up in a situation where obligations are imposed on researchers without funds to support it. Everybody is working on their report this week. They've asked us to send our executive summary, so we'll do that.

Vicki: "There's one punchy sentence missing at the end of the summary: the idea that there's a limit of what we can do, driven by funding and specific needs. If you want to use the data, there needs to be a more thoughtful effort on how you fund it. It's in the summary, but it's too loose, not crisp enough. You need to get that punch at the end: one or two sentences to pull it together.

Tom: "I was about to say just that. That's what the report needs. We can do it in real time or make our best effort to fine tune it and send around by email.

Roscoe: "We need quorum of ASCAC to agree to forward the report with revisions. We can then go back for editing and circulate final draft.

Tom: "I think it's just the right thing to do."

Roscoe: "You move to accept the report. There is a process."

Marsha: "So you need a motion to accept the report?"

Vicki: "I'm fine with revisions, just with a couple of sentences added."

Roscoe: "That's what I want to be sure we have in the notes, to go back over it?

(Unidentified man:) "Roscoe, where do you see this heading? What is the Office of Science planning on doing? The different offices will cross-talk?

Roscoe: "I don't know. This is in response to a law, part of the COMPETES reauthorization, so they will have to make a response to that. The Office [of Science] is fact-finding. …We're in middle of a motion. (Susan seconded) Is there discussion of that motion? Are you ready to vote? Sounds like we are; we can certainly take more time. [pause] ASCAC members, we'll do a round, one by one (say aye).

Roscoe called out participants' names:

"Marsha: yes, approve.
Jim: yeah (…motion subject to revision)
Vivek: yes
Susan: yes
Tom: yes
Linda: yes
Jackie: yes
John: yes
Vickie: yes
[pause]
[Roscoe:]   yes.

Bill Tang: "I had trouble getting on at the beginning, but I vote yes."

Roscoe: "I failed to write you down. So thank you.

Roscoe: "I'm not sure I need to go through that long bulleted list. I don't want to go through it in great detail. We can look at the items to suggest if there's anything else we might say.

Who makes determination? It's on a local level. There's the issue of software: how access is provided/controlled/limited in any way, [software that] comes with additional functionality. [There was discussion of the merge of written data versions.(?)] Whether peer review is a condition of dissemination.

DOE is the body by which policy is currently upheld? Is this accounted for by existing policy or practice?

Roscoe: "I don't know if that suggests additional things to mention. We don't need to go into detail for each bullet.

Jim: "I'd be happy to take suggestions from everyone individually. Starting with your comment to type this up with more punch. On these bullet items, if some need additional fleshing out in the current layout, the way it's organized, if you send suggestions, I'll get them into the report.

Roscoe: "Let's put a deadline on it."

Jim: "As soon as possible. Within an hour or two of this telephone call, while it's fresh would be good. [The input] doesn't need detail. If there are other areas to reference, it's remarkable if there is any policy [to be referenced] out there. We can include policies driving dissemination practices; that would be great.

Marsha: "I read a draft report on what the community is doing for a journal committee, but it has a section on what journals' practices are on archiving, and how that's evolving. Might be something to reference how things that are evolving. Tom, when will that be out?

Tom: "I don't know."

Marsha: "I'll ask Randy and send an email.

Roscoe: "…For this AIP and publishers' letter: see if we can refer to it instead of attach it to our report.

Jim: "I didn't know how to do that either. He presumably has that report.

John: "I suggest we have a deadline by Friday night for anything to include. So he has the weekend knowing he doesn't have additional things to deal with.

Jim: "I'll turn around another draft. Christine: is it all right if I send it directly to the committee?

Vicki: "I will send a few sentences of what I think would be punchy.

Roscoe: "If there aren't additional things to bring up… [pause]   We're finished.