Minutes
Biological and Environmental Research Advisory Committee
Hilton Hotel and Executive Conference Center
Rockville, Md.
February 21-22, 2013

BERAC Members Present

Gary Stacey, Chair
Dennis Baldocchi
Janet Braam
Judith Curry
James Ehleringer
Susan Hubbard
Andrzej Joachimiak
L. Ruby Leung
Gerald Mace
Sabeeha Merchant
Joyce Penner
Karin Remington
G. Philip Robertson
Martha Schlicher
Jacqueline Shanks
Gus Shaver
David Stahl
Judy Wall
Warren Washington
Minghua Zhang
Huimin Zhao

BERAC Members Absent

David Randall
James Randerson
William Schlessinger
Herman Shugart

About 55 others were in attendance during the course of the two-day meeting.

Thursday, February 21, 2013
Morning Session

Before the meeting, the Committee members received their annual ethics briefing by **David Krentel** of the DOE General Counsel's Office.

The meeting was called to order by the Chair, **Gary Stacey**, at 9:00 a.m. He welcomed the members and visitors and pointed out that the meeting was being recorded and streamed live to the web. He asked the members to introduce themselves.

After the member introductions, **Sharlene Weatherwax** was asked to provide an update on the activities of the Office of Biological and Environmental Research (BER).

The BER budget is broken into two categories, the Biological Systems Science Division (BSSD) and the Climate and Environmental Sciences Division (CESD); it finished up FY12 with $609.5 million. BER is still talking about its FY13 request and is operating under a 6-month continuing-resolution allocation at the House mark, the lower of the Senate and House marks. It has been allocated $264. 3 million for the first 6 months of the fiscal year. No new projects or budget restructurings are allowed; BER is not affected in that regard.

The outgoing members were thanked for their service to BERAC: James Tiedje, Raymond Wildung, Gary Sayler, and Gregory Petsko.

In the Office, Arthur Katz retired after 30.5 years, and Sally McFarlane has joined the staff as an Atmospheric Scientist. A number of active recruitments are in progress for a microbiologist, a science assistant for each Division, a data informatics program manager, and a program manager for the bioenergy research centers.

BERAC was charged to look at BER's facilities and ensure that these investments are being optimized, are aligned with portfolio priorities, engage the research community, and move science forward. Not much had been put into the long-range plans about facilities. The facility needs are to be assessed in terms of the facility's ability to contribute to world-leading science, its readiness for construction, and its estimated construction and operation costs. Only items exceeding $100 million total project cost were to be considered. A short letter report is due March 22, 2013. The facility science is to be binned, and the construction is to be assigned a status. This assessment started by looking at the BER long-term vision report, which called for looking at BER science from a systems perspective, and the BER technology innovation report, which focused on understanding data and models.

BER wants to support experimentation and modeling that can reliably predict outcomes and behaviors of complex biological, environmental, and climate systems, enabling effective and innovative solutions for DOE missions and strategic goals. BER recognizes that a multidisciplinary effort is required to conduct the complex systems science that is emerging as the only way to gain broad insights. It is hoped that scientists will increasingly propose research by combining data from disparate sources.

In support of the facilities charge, the Office gave the Committee information on all BER user facilities. It pointed out that virtual laboratories that are networked but not collocated are possible. What has been proposed is a BER User Facility for Biological, Climate, and Environmental Data, Analysis and Visualization.

Stacey noted that, initially, it may be three to five models; the National Science Foundation's iPlant collaborative had similar goals. He asked how Weatherwax saw the facility being implemented. She replied that the Office understands where iPlant is going. But having individual targeted campaigns goes against the centralization of the Joint Genome Institute (JGI). But this is not an either/or situation. The facility is not to be just a data warehouse. An interactive environment is desired. Interaction among disciplines is sought. The holistic view is what people are excited about. He feared that the idea of integration will be lost. The iPlant initiative was isolated from the researchers who had started the idea. Weatherwax agreed that the people who produce the data must be integrated with those who do the modeling. Data management and computer infrastructure must be focused on and funded.

Joachimiak commented that one also wants to set up an experimental facility that can capture a lot of data and integrate those data into a knowledgebase. Weatherwax responded that that is what was done with the Knowledgebase (Kbase). This process can be fleshed out into additional disciplines. A flexible system is needed that can handle different/new data types.

Washington noted that there was nothing that showed partnerships with other offices of the Office of Science (SC). Weatherwax noted that BER has partnerships with the Office of Advanced Scientific Computing Research (ASCR). The Office recognizes the need for such partnerships and the need to build a budget case to be able to support them. Mace observed that Cyberinfrastructure, Analytics, Simulation, and Knowledge Discovery (CASK) and integrated virtual field laboratories need to go together.

BER funds a large amount of research through university grants and cooperative agreements and through national laboratory science focus areas (SFAs). What was intended with the SFAs was to encourage, facilitate, and effectively manage integrative and collaborative programs at the national laboratories to achieve scientific research and solutions of the highest quality in support of BER strategic goals. It was desired for the work at the national laboratories to be complementary to research conducted at other institutions, such as universities or the private sector. The national laboratories are also expected to develop and evolve their research programs over time to identify, build, and anticipate new areas of science and future research needs and challenges. Additionally, as BER's strategic goals change and as science progresses, the national laboratories are expected to reconfigure SFA programs to meet these changing research needs. BER now has a number of SFAs, some based on long-established programs and some new. An attempt is being made to align the SFAs with the Office's budget categories.

Both divisions have completed their initial cycles of establishing, reviewing, and adjusting SFAs; this process will continue. BSSD will be developing a strategic plan this coming year and will adjust SFAs as necessary and appropriate. CESD will use their new strategic plan to transition their SFAs so that the majority of nonfacility funding to the national laboratories is aligned with an SFA. BER sent a letter to laboratory directors to update them on SFA definitions, management, transitions, reviews, and collaborations.

The FY13 budget is about evenly split between the two divisions. There is also a healthy balance (about 2 to 1) between R&D and facilities.

Two recent highlights from BER-funded research include, research on the genetic basis for bacterial mercury methylation at Oak Ridge National Laboratory and the work on marginal lands as a valuable resource for sustainable bioenergy production at the Great Lakes Bioenergy Research Center (GLBRC).

Another previously funded, long-term BER program, the development of an artificial retina, was recently in the news as the device developed received Food and Drug Administration (FDA) approval; this effort took 10 years and $75 million. For other long-term research efforts, BER-funded scientists Sally Chisholm, Lucy Shapiro, and Lee Hood received the 2011 National Medal of Science from the President. Other awards to the BER community include Sabeeha Merchant's (BERAC) being elected to the National Academy of Sciences, Tim Donohue's (GLBRC Director) being elected President of the American Society for Microbiology, and Wanda Ferrell's (BER Program Manager) receiving the prestigious "Cleveland Abbe Award for Distinguished Service to Atmospheric Sciences" from the American Meteorological Society and being made a fellow of that society.

The Department of Energy (DOE) and the National Oceanic and Atmospheric Administration (NOAA) signed a memorandum of understanding (MOU) in late January to more efficiently develop an understanding of the climate system and science-based prediction tools in support of the nation's needs for secure energy, environment, water, food, health, and economic well-being.

Penner commented that the MOU was interesting and asked how NOAA and DOE will work together. Weatherwax replied that having a broad MOU will allow flexibility for cooperation between the agencies.

Todd Anderson was introduced to give an update on the BSSD.

SFA triennial reviews in FY13 will be held at Pacific Northwest National Laboratory in biofuels research and at Lawrence Berkeley National Laboratory (LBNL) and Oak Ridge National Laboratory (ORNL) in radiochemistry and imaging. The 2013 Genomic Science Program Annual Principal Investigator (PI) Meeting will be held at the Bethesda North Marriott Hotel on February 25-27, 2013.

New funding notices have been posted on Plant Feedstocks Genomics for Bioenergy: A Joint Research Funding Opportunity Announcement of the U.S. Department of Agriculture (USDA) and DOE (full applications are due Feb. 25, 2013) and on Systems Biology Enabled Research on the Role of Microbial Communities in Carbon Cycling (pre-applications are due on Mar. 4, 2013, and full applications on Apr. 19, 2013). Multidisciplinary research is encouraged.

A new user facility-based research opportunity is the first-ever joint call between the Environmental Molecular Sciences Laboratory (EMSL) and JGI, which is focused on plant, fungal, soil, and microbial interactions and physiology. Proposals must require capabilities from both facilities. Letters of intent are due between Feb. 11 and Apr. 8; invited full proposals are due May 27; and approved proposals start Oct. 1. Stacey commented that such collaboration is wonderful. He predicted great things from this collaboration. It should stand as a model for future research activities.

The Office is also excited about the Knowledge Base (KBase) rollout in February 2013 at the Genomic Science PI meeting. Through a series of talks, demonstrations, and tutorials, the KBase team will illustrate the early functionality of KBase and demonstrate the capability to integrate your own research methods into the KBase environment. A long-term challenge for KBase is to maintain connections with researchers.

The FY12 Reviews have been completed for the BioEnergy Science Center at Oak Ridge National Lab.; Great Lakes Bioenergy Research Center in Madison, Wisc.; and Joint BioEnergy Institute at Lawrence Berkeley National Lab. The reviewers' comments have been compiled and sent to the facility managers. The Bioenergy Research Centers (BRCs) reviews showed significant progress toward understanding sustainability of bioenergy crop production; plant metabolism and techniques to decrease biomass recalcitrance; pretreatment methods to increase the efficiency of cellulose extraction; enhanced enzymatic methods to produce sugars from cellulose; and modifications to microorganisms to combine conversion capabilities, to tolerate biofuel-production conditions, and to produce a range of biofuel compounds. During the past 5 years, the BRCs have produced 1100 journal articles, 286 invention disclosures, and146 patent applications.

A couple of BSSD science achievements were highlighted:

- From the BioEnergy Science Center, Nanoscale Architecture of Plant Cell Walls Determines Their Accessibility and Digestibility by Enzymes revealed different mechanisms of enzymatic breakdown of biomass by fungi and large cellulosomes.
- From the Joint BioEnergy Institute, Making the Best Biofuel-Producing Microbes Identify Themselves facilitates the selection of microbial strains that produce large quantities of any small molecule, an important step toward the development of renewable biofuels.

- From the Genomic Science Program, A More Efficient Approach to Map Biochemical Reactions at the Atomic Level achieved very high computational efficiency and has many applications across scientific disciplines.
- From Oak Ridge National Laboratory and The University of Tennessee, Discovery of New Types of Nitrous-Oxide-Consuming Bacteria identified a novel metabolic pathway for N_2O consumption in the soil bacterium *Anaeromyxobacter dehalogens,* perhaps representing the missing sink for N_2O in soil ecosystems.
- From Lawrence Berkeley National Laboratory, Small-Angle X-ray Scattering Study of Motor Assembly and Motility in *Archaea* revealed detailed and global FlaI activities in transducing nucleotide binding and hydrolysis into translational and rotatory motions suitable for assembly and motility without destabilizing the integrity of the hexameric crown assembly.
- From the University of Wisconsin, Low-Dose Radiation-Induced Epigenetic Alterations in the A^{vy} Yellow Agouti Mouse Model showed that low-dose radiation can elicit an epigenetic spots and that those low-dose-induced epigenetic changes play a role in radiation hormesis.

JGI has produced a number of significant publications.

A break was declared at 10:56 a.m. The meeting was called back into session at 11:16 a.m.

Gary Geernaert was asked to present an update on the CESD.

In staffing the CESD, a Science Assistant and a Physical Scientist are being sought.

The Division's three platforms of observational infrastructure, community models, and community data infrastructure are being integrated; some components still need to be integrated.

Executing the Division's strategic plan calls for (1) accelerating capabilities in predictive modeling (a workshop was held in Berkeley to look at gaps in knowledge to aid this activity); (2) observational and data capabilities in ARM, EMSL, and data management (a growing need); (3) MODEX (analysis-based integration of "MODeling and EXperiment"), which was supported by several workshops last year; (4) balanced funding mechanisms (SFAs, boutiques, and university grants); and (5) interagency collaboration [with NOAA, the National Science Foundation (NSF), and the National Center for Atmospheric Research (NCAR)].

Currently, about half of the Division's funding goes to universities and half to national laboratories. More and more of the long-term assets will be shifted into the SFAs, with about 75% going there by 2016.

Recent CESD events include PI meetings and meetings/workshops on the water cycle, Interagency Group on Integrative Modeling (IGIM), Coupled Model Intercomparison Project (CMIP), the Atmospheric Radiation Measurement/European Union (ARM/EU) joint meeting, model-development strategy, aerosol chemistry, and North American Carbon Program.

Three funding opportunity announcements in FY13 for Terrestrial Ecosystem Science (TES, for which proposals are under review), Atmospheric System Research (ASR, will be issued in March and due in May), and Green Ocean Amazon (GOAmazon, which will be issued in March and due in June). In addition, there is also a National Aeronautics and Space Administration /

Research Opportunities in Space and Earth Sciences (NASA/ROSES) call for a multi-agency effort that includes DOE.

Triennial SFA reviews will be held for in the spring and fall for the LBNL TES, LBNL ASR, LBNL Climate, Brookhaven National Laboratory (BNL) ASR, and Pacific Northwest National Laboratory (PNNL) Subsurface Biogeochemical Research (SBR).

The ARM Climate Research Facility’s current activities include the first ARM Mobile Facility (AMF1; FY13), the Marine ARM GPCI Investigation of Clouds (MAGIC) [GPIC is the Global Energy and Water Cycle Experiment–Cloud System Studies Pacific Cross-Section Intercomparison], in which the Horizon Spirit makes a round trip from Los Angeles to Honolulu each 2 weeks; AMF2 (FY12-13), the Two Column Aerosol Project (TCAP) on Cape Cod; and AMF3 (FY13 and beyond), in which unmanned aerial vehicles were demonstrated in November 2012 in Oliktok, Alaska. Also coming up are airborne campaigns in FY13; GOAmazon from January 2014 through mid-2015; and a study of biogenic aerosols and climate in January 2014 in Hyytiala, Finland.

New EMSL capabilities coming online in FY13 include the Radiochemistry Annex and a next-generation supercomputer.

Several science highlights of CESD-sponsored research were cited:

- For more than 30 years, scientists have debated whether the most recent major cold period on Earth was triggered by meltwater from the Arctic or from the Gulf of St. Lawrence. Research at the University of Massachusetts indicates that only meltwater from the Arctic significantly weakens the meridional overturning circulation (MOC), pointing to the Arctic’s MacKenzie Valley as a potential trigger for climate change.
- Multiple effects of aerosol pollution on the South Asian monsoons have been observed. Research at PNNL found that a slow response that cools the sea surface, weakens the north-south temperature gradient, reduces the Hadley circulation, and decreases the northward transport of moisture over the continent is dominant over the fast response that produces an east-west asymmetry of circulation responses.
- A new treatment of small-scale variability at PNNL significantly increases the simulated amount of fair-weather clouds over the central United States, consistent with observations. The net decrease in the amount of model-simulated sunshine reaching the surface is also consistent with observations.
- The new, physically based Model for Scale-Adaptive River Transport (MOSART) for runoff-routing modeling was developed at PNNL. It provides a flexible framework for modeling terrestrial fluxes into the ocean for complete linkages across the atmosphere, land, and ocean in Earth-system models and serves as a cornerstone for integrating the human- and Earth-system components of the water cycle.
- A Princeton University simulation of insect disturbance on carbon dynamics in the New Jersey pine barrens showed that gypsy moths significantly affect the carbon cycling and carbon balance of forests and need to be incorporated into forest-development models. It also showed how thresholds of stress can affect forests.
- A study at the University of Missouri of the proteomics of soybean-root-hair interactions with bacteria gave a new understanding of the interactions of root hairs with bacteria that may help redesign plants and improve crop yields.

Stahl asked, on MODEX, which research example would be the best example of integration. Geernaert replied that it would be the Next Generation Ecosystem Experiment (NGEE), which takes recent fruits of modeling, looks for gaps, and integrates experiments into that framework to address those gaps.

Baldocchi asked what role DOE would be interested in with the Integrated Common Observation System (ICOS). Geernaert responded that the Office has had discussions about NEON and CDIAC but has not yet discussed ICOS. It wants to coordinate data-management plans and integrate infrastructure. Baldocchi followed up by asking what DOE's role/position was in the fracking debate. Geernaert answered that fracking is not something that BER gets involved in. It is an issue for the Office of Fossil Energy.

Judy Wall gave a science talk on the genetic basis for bacterial mercury methylation.

Ionic mercury in soils and sediments is converted to methyl mercury, a neurotoxin that bioaccumulates in the food chain. It causes birth defects and disorders of the brain, reproductive system, immune system, kidney, and liver at extremely low levels in food. It has caused widespread public health disasters in Minamata, Japan, and in Iraq.

Mercury has both natural and anthropogenic sources. The largest source is now coal-burning power plants. The atmospheric deposition of mercury has increased significantly during the industrial age. The only long-term storage site for mercury is sediment.

Methylation can occur by radical, anionic, and cobalamin pathways. Cofactors that transfer methyl groups include S-adenosylmethionine, N^5-methyltetrahydrofolate, and methylcobalamin. The source of the methyl group in methyl mercury has been found to be methyl tetrahydrofuran, and a cobalamin-dependent (i.e., corrinoid) protein has been found to be involved, implicating the reductive acetyl-CoA biochemical pathway. However, the protein was not characterized or purified because unexplained inconsistencies and technical difficulties delayed progress.

The phylogenetic relationships among mercury-methylating delta proteobacteria were defined, and the necessary chemistry was determined. A methylate cofactor [CH_3-Co(III)-protein] is needed, a methyl group is transferred to Hg^{++} by transferring electrons, and two electrons are needed to regenerate the Co(I)-protein from the Co(III)-protein cofactor. The methyl group was found to transfer in the reductive acetyl-CoA pathway with two CH_3^+ transfers. The organism had to have a gene encoding a CFeSP-like protein that is unique to the three confirmed methylators among the *Desulfovibrio.* All unique genes were queried, and 36 genomes were identified. A fairly good fit of the HgsA protein was found.

Comparative genomics was employed, and ferredoxins were identified as likely methylators. A mercury-methylation pathway was proposed, and it was desired to confirm the two genes, hgcA and hgcB (hgcAB), as being necessary for methylation of mercury. Marker-exchange deletions of hgcAB were used in ND132, and antibiobic resistance (Kan^R) was shown to disable the methylation. Complementation of deletions (replacement of the genes) was used to confirm the methylation capability. So far, a perfect correlation has been seen between the presence of hgcAB and the ability to methylate mercury: both hgcA and hgcB are essential for methylation but are not sufficient on their own; something else is needed. Microbes with complete sequences are now being tested.

The question arises about what the role of the C-terminal domain is in hgcA. It seems to be membrane-associated. The transmembrane domain of hgcA was deleted to see the effect on activity. The deletion reduced activity below current detection levels. The conclusion is that the transmembrane domain is required for methylation by ND132 hgcA. The implications of this conclusion are that:

- How mercury research is performed will change globally.
- The proposed mechanism is *new chemistry*, which, if confirmed, will open up new areas of research.
- There is now a potential biomarker for methyl mercury production.
- Correlating gene, protein, and organism abundances with methyl mercury formation rates and yields will lead to improved and more sensitive biogeochemical models.

Many questions still remain regarding the molecular mechanism of mercury methylation, such as what is the methyl donor; what are the reaction mechanism, function, and structure; and whether the hgcB is the only donor of electrons.

Merchant asked what the membrane-domain requirements were. Wall said that she did not know; there may be a transport function.

Robertson asked where the reverse methylation reaction occurs. Wall replied that it has been observed and reported but is a difficult assay to do. One can follow the two reactions and document the demethylation. They are not demethylated in the way that methyl organomercury is in the *mer* operon; it is a different property. Researchers are still struggling to figure out how that actually happens, whether it is activated demethylation or reductive demethtylation.

Stahl asked how those genes relate to resistance of the bacteria to mercury poisoning. Wall answered that the sulfate reducers are robust to mercury challenge; they can bind up the mercury. The general pattern of growth allows them to resist the toxicity of mercury. One can look at making them more or less tolerant.

A break for lunch was declared at 12:41 p.m.

Thursday, February 21, 2013

Afternoon Session

The meeting was called back into session at 2:16 p.m. Stacey announced that a subcommittee had been formed to respond to the charge on facilities from William Brinkman, Director of SC. All SC advisory committees received similar charges. The Subcommittee has held three teleconferences, made recommendations, and published a draft report. He opened the floor to discussion of the draft report.

Hubbard asked for the definition of a facility. Weatherwax said that the Office interpreted it as the definition of SC user facilities.

Robertson said that being given this task was opportune because new facilities will be needed, and the pieces of this recommended plan needed to be dealt with in a coordinated fashion.

Mace stated that it is just one facility that is being talked about with three components: the Integrated Field Laboratories; the Biosystems Frontier Network; and CASK. Remington added that it ties in visualization with other data components. Mace said that it also involves community models, and they talk to each other across time and space. Visualization is on top of all that. The integration is what it is all about.

Joachimiak noted that what is being proposed only makes sense if all three components are there. It is desirable to carry out the activities in real time. DOE has great computing capabilities. The power of the proposal is the integration of all the components.

Merchant stated that it is important to have a pilot project to iron out the common goals.

Shaver was confused about the discussion and where it was going: the importance of integration. The development of the parts is crucial. He asked what the intent of the letter to the Committee was. Stacey responded that it was based on the CASK idea. Throughout the discussion, a holistic approach was assumed. Shaver reiterated, the letter pointed to CASK first and then to integration.

Ehleringer noted that, when one thinks of a facility, one thinks of a building. This work needs to be distributed. There needs to be some way to reflect how all this work is going to be pulled together. Stacey responded that BER is the natural home of nonmedical systems biology in the federal government. He was thinking of focusing the letter on systems science as the science of the 21st century, of which CASK is the piece that is missing. (20th century science was the study of molecules; 21st century science is how molecules interact.) Leung suggested pointing to the vertical integration of information. Zhang said that one must include the field laboratories, the collection of data, and the ability to access and analyze the data.

Shaver pointed out that this is not a new idea. The integration of modeling and experimental studies has been talked about since at least 1972. This is what everyone has wanted for a long time. This report emphasizes that it is a new idea, which it is not. Stacey said that it does not need to be a *new* idea.

Baldocchi stated that researchers are working on a great range of scales in complex systems. Based on that, one needs to integrate what is everywhere all the time.

Stacey asked how prescriptive this letter needs to be. Does it need to call for five pilot projects? It is assumed that community meetings will be needed for input. Weatherwax replied that what Brinkman wants is backup information for a Critical Decision Zero (CD-0), the Justification for Mission Need for a new project.

Penner suggested that, being limited to two pages, one might think about *how* this laboratory would work: a researcher sitting at a desktop in a university accessing models, simulations, and data from disparate sources.

Hubbard asked if the concepts of "virtual laboratory" and "CASK" should be put up front. Robertson pointed out that the engineering problems for the CASK are more mature than those for the other two components. Stacey said that the letter needs to address (1) how the recommendations address the production of world science and (2) the readiness of design. It needs to review current facilities and the facilities of the future. No one wants to downgrade the current BER facilities.

Washington said that each community has its own formats. A system needs to be built to interface those formats. Careful attention should be given to integration by translating one format to another.

Mace pointed out that energy exploration and production is expanding into the Arctic and that the environment is not the only issue to be addressed. Hubbard stated that these other issues need to be tackled but probably in the implementation stage.

Stacey recalled that Judy Wall mentioned trying to annotate the unknown sections of the genome. Remington said that the Virtual Laboratory should provide an opportunity for a collaborative meeting space so researchers could share tools and work together. Stahl asked what the glue was that held these components together. Such an operational question is important to address. Stacey agreed that there is a problem of comparability of data and ventured that perhaps this Virtual Laboratory will be a step forward in being able to make many measurements simultaneously. Wall added that one needs expertise on, say, entomology. If there were a network of experts, it would allow a student to go to the entomology expert on the rare occurrence when that expertise was needed. Being able to study many samples simultaneously is a huge problem for many disciplines. Braam pointed out that there may be other ways to get the needed information. Wall agreed; there needs to be flexibility to accommodate alternative approaches. Joachimiak said that what is being talked about is collecting huge amounts of data and making sense out of it. Another option is to develop new tools to do the analysis, as in crystallography. Such capabilities need to be opened up to other sciences.

Curry said that to go from where we are to where we want to be is overwhelming. The community is being asked to put all its eggs in one basket. One must be careful in making that choice. Even the near-term steps are daunting. The Committee needs to be sure that what is recommended is what will solve the problems that need to be solved. This is a high-level strategic objective.

Hubbard asked if there were a sense of the timescale that should be looked at.

Mace said that the community kind of knows where it wants to go but does not know how to get there. If the virtual facility comes together, it will be a quantum change in the community's ability to understand.

Shanks stated that there is a major fear: the computation component depends on both of the other components. An entity is needed that can integrate all the components.

Stacey said that one issue is the programming and visualization tools. Visualization tools do not always play well with other tools. Widely adopted tools are those that play well with others. Writing codes that address the problem at hand really well often produces codes that do not work with other codes. Remington suggested that what is needed is computer scientists who specialize in integration, not tools. Hubbard pointed out that heterogeneous data is another problem that needs to be addressed. Remington said that the problem of heterogeneous data is included in the federal Big-Data Initiative. Developing and maintaining a web interface is a terrible challenge.

Merchant stated that data are collected in different formats by different people and at different times. Data can become inaccessible because of this. Pilot projects would bring this problem to light.

Schlicher said that this conversation reflects discussions that researchers at Monsanto have every day. More transparency (i.e., data sharing) is needed between the private and public sectors. The discussion of the lack of uniformity and standards should not stop at the boundary between the public and private sectors.

Zhang pointed out that the challenge is to define a facility that no one else has. Stacey said that BER focuses on climate and sees microbes as a portion of that problem. The Committee needs some help thinking about that. Mace said that this project is what people have wanted to do since they began modeling. If DOE invested $100 million, one could do things that one never could do before.

Baldocchi said that what would help would be a map that one could zoom in or out on and see many levels of data.

Stacey asked, what would be an idea that would be as exciting as the human genome? Ehleringer offered understanding climate change in a fundamental way. Shaver suggested describing the Earth as a single system. Facilities would have to be created to do this. The people who invented systems science were engineers and economists. One should build on what they have done. This type of facility is needed to address this complex system.

Leung pointed out that there are many types of models. It would be helpful to compare models and to compare their results to experimental results. Merchant added that it would be good to inter-relate different types of data for the same sample. DOE has great user facilities to make measurements. Joachimiak stated that community has been trying to develop predictive models. This is a good idea, and one should stick with it.

Schlicher observed that a number of agencies would fit under the Google Earth framework. The Committee should promote it as such.

Braam noted that the KBase information will allow predicted consequences of perturbations, which is very powerful. Mace added that it is based on reality, also; model results are not. Stacey stated that Google Earth was mentioned in the original letter, but it may not belong in a two-page letter. It is an example of how to look at environmental data. Robinson liked the idea of using the analogy to make the subject matter understandable to someone like the President.

Stacey asked the Committee members to think about (1) whether something was left out, (2) the logistics of getting the letter written, and (3) words that would capture one's imagination.

A break was declared at 3:43 p.m. The meeting was reconvened at 4:00 p.m.

Stan Wullschleger was asked to discuss the early results from the Next-Generation Ecosystem Experiments (NGEE) Arctic Project.

The goal of the NGEE is to deliver over 3 years a process-rich ecosystem model, extending from bedrock to the top of the vegetative canopy, in which the evolution of arctic ecosystems in a changing climate can be modeled at the scale of a high-resolution Earth-System Model grid cell.

A 2012 report pointed out the ocean, atmosphere, and land processes that are important to climate modeling. Many other reports detail ecosystems that would be good to study and model. Arctic ecosystems contribute to climate feedbacks because of their large land area with permafrost containing about 1700 Pg of carbon, of which up to 90% could be lost to the atmosphere by 2100. The active layer thickness is increasing. Model-based projections of permafrost vulnerability and potential carbon loss associated with climate warming indicate that 150 to 450 Pg of carbon could be lost to the atmosphere. There was a lot of variability in the estimates, largely because of the representation of ecosystem processes. It is difficult to organize experiments and observations into model representations.

A community workshop was held in 2010, and it was followed by other gatherings, resulting in an overarching science question (How does permafrost thaw and degrade, and how do the associated changes in landscape evolution, hydrology, soil biogeochemical processes, and plant community succession affect feedbacks to the climate system?) and some directions to pursue.

Surface–subsurface interactions lead to consequences for landscape evolution. It is desirable to characterize the North Slope of Alaska to determine the climate effects of these landscape changes. The Barrow Environmental Observatory (BEO) is 7475 acres set aside for scientific research. It is adjacent to the ARM Arctic Facility, which gives a good deal of pre-existing climate and weather data. The area has drained thaw-lake basins, thaw lakes, and interstitials (e.g., polygons). Intensive field sites have been established. The topology rises from 4 to 7 m above sea level, but even that small variation strongly affects water and nutrient distribution across the landscape.

Early results include 18 publications, 56 abstracts or presentations, 13 conferences and workshops attended, and 7 storylines and press releases.

Geophysical techniques were used to describe the landscape, understand processes, and provide modeling input. Cluster analysis revealed consistent geomorphic and subsurface zonation. This is a new approach for quantifying co-variants of land and subsurface processes in high resolution and in a minimally invasive manner. Each zone had a unique distribution of hydrothermal-geochemical properties. Soil cores were used to study the spatial variability of bacterial composition in high and low polygons. Ground-penetrating radar was used to probe soil moisture content. These data were analyzed with Bayesian approaches. Electrical-resistance tomography was used for ice-wedge characterization. One can return to the sites and perform appropriate follow-up activities when an observation is made that indicates that something is going on.

Most Earth-system models (ESMs) use the Farquhar et al. approach to estimate gross primary production (GPP). One can use A/Ci curves (carbon assimilation as a function of the intercellular concentration of carbon dioxide) to model ecosystem production.

Synthesis efforts to derive estimates of the maximum rate of Rubisco-mediated carboxylation (V_{cmax}) for arctic plant species have historically not contained enough data to generate parameters for the Farquhar et al. model of photosynthesis. The NGEE Arctic Project is beginning to provide field-relevant estimates of photosynthetic biochemistry. Some of the data collected were for V_{cmax} and leaf nitrogen content. Those data allow comparison of models' representations of arctic ecosystems.

A methane module was sought for such models as the Community Land Model. These modules are highly sensitive to differences in temperature- and oxygen-response factors. Extrapolating parameters measured in temperate soils introduces significant uncertainty in predictions. So, 27 permafrost cores were drilled from Barrow polygonal tundra and analyzed with X-ray computer tomography. Frozen cores were processed for anoxic incubations and soil organic matter (SOM) analysis. Carbon release from the permafrost layer was significantly greater than that from the mineral horizon.

There are microbial-abundance variations above and below the permafrost transition. Collaborative analysis of microbial communities and SOM chemistry will improve models of decomposition. In process-level modeling, the measurements are, at best, at the plot level; 10-

km^2 areas need to be prepared to use these data in climate models. One needs to go from the fine scale to the intermediate scale to the climate scale to predict future arctic-ecosystem responses to climate change.

Los Alamos National Laboratory is developing the Arctic Terrestrial Simulator for a single polygon domain (25 m × 25 m) to a multiple-polygon domain (100 m × 100 m) to represent arctic vegetation response. Modeling ice wedges is also of interest.

The banded graph cut (BGC) algorithms in the Community Land-Surface Model (CLM) are being implemented to PFLOTRAN, a massively parallel subsurface flow and reactive-transport code, with CLM's subsurface routines being replaced by PFLOTRAN but its surface-flux algorithms and implementation being retained. The current prototype interface between CLM and PFLOTRAN is being refined.

In pursuing the NGEE Arctic goal of developing a process-rich land-surface model for improved climate prediction, new parameters and algorithms, initialization data sets, evaluation data sets, and discovery science have been and continue to be looked at. Collaborators are welcome for sample analysis or field studies.

Shaver asked about upscaling in space and time. Barrow is not like a lot of the Arctic, and one cannot predict how the landscape will change with time. Wullschleger agreed that these systems are complicated and interconnected. An effort is being made to try to see how these regions vary, see how characteristics vary from region to region, and then use those differences to scale across regions. This project is not just about carbon but also about nitrogen. It is hoped to move in that direction. The project is looking at cores' responses to temperatures. There is other work going on that contributes to time scaling.

Robertson asked how the ARM facility was being used or interacted with. Weatherwax said that the facility provides radiation and climate inputs for Barrow, Alaska. Ferrell added that a meteorological tower is there to measure carbon dioxide production and flux. Other opportunities for interpretation will emerge as one looks at radiation balance etc.

Baldocchi asked if there were any vision to go to year-round operation. Wullschleger replied that there is the hope to have year-round eddy covariance data collection. The team will be there from April 15 to November 15. One does not want to miss shoulder events, so year-round operations are important.

Shaver asked if there were any plans to go back to do observations of above- and below-ground warming. Wullschleger pointed out that there were cost, environmental, and logistics limits, so the team has backed off from that goal; however, the door is not closed on it. Some such observations could be done on a small scale.

The floor was opened to public comment. **Reinhold Mann** said that some boundary conditions should be set to the facility discussion, and the review process should be discussed. **Rick Stevens** said that the key observation that just integrating different types of data will open insights can be turned around, and articulations of three or four specific key questions would support the budgetary hearings better.

There being no other comments, the meeting was adjourned for the day at 5:10 p.m.

Friday, February 22, 2013

The meeting was reconvened at 8:30 a.m. **Renu Joseph** was introduced to report on the Workshop on Community Modeling and Long-Term Predictions of the Integrated Water Cycle.

The Workshop was held on September 24-26, 2012, in Washington, D.C., to understand processes and feedbacks in the climate system. The workshop goal was to identify challenges of next-generation human–Earth-system models for improving long-term predictions of the regional-scale integrated water cycle. There were 80 invited participants, including representatives from eight agencies. It culminated with an interagency panel discussion.

Water is essential for energy systems, ecosystem services, and a wide range of life-sustaining and other critical human activities. DOE has an interest in the Community Earth System Model (CESM) and the Global Change Assessment Model. The subsurface water components are not well represented in these models. The integrated water cycle consists of natural processes and human systems. These processes occur at different (but overlapping) scales.

The workshop addressed six main topics:

- Multiscale behaviors of the water cycle
- Human–Earth-system interactions and their impacts on the water cycle
- Challenges for land surface/hydrologic modeling
- Model testing, analysis, and evaluation and data needs
- Prediction, analysis, and uncertainty quantification of water-cycle mean and extremes
- Use-inspired water-cycle research to meet the most pressing energy and environmental challenges

Three science grand challenges were identified and, subsequently, the integrated-modeling experiments that will be needed. The first science grand challenge is modeling the multiscale atmospheric and terrestrial processes and interactions. It requires understanding the scaling and scale interactions of atmospheric industrial processes; representing the multiscale processes and interactions across systems in Earth-system models; and acquiring a model testbed, evaluation, and data.

The second grand challenge focuses on the integrated human–Earth system and its links with water resources. It requires an understanding of the roles of human systems at different spatial and temporal scales in the coupled system; representing a wide range of human–Earth-system interactions across scales; acquiring a model testbed, evaluation, and data; and advancing and validating the understanding of the role of human–Earth interactions in water-cycle changes.

The third science grand challenge deals with advancing prediction and uncertainty quantification for decision support and mission-oriented objectives. It requires advancing model predictions; developing uncertainty quantification, metrics, and observations; and developing a team approach to use-inspired research.

The first integrated-modeling experiment addresses the implications of land cover and land-use change for regional climate, water resources, and energy pathways in the United States.

The second integrated-modeling experiment addresses multi-model hierarchies to address a wide range of user needs for predicting the regional integrated water cycle to evaluate the predictive skill of the model.

The third integrated-modeling experiment addresses the sustainability of water and energy resources in the eastern versus western United States under the stress of climatic and societal changes. In the east, runoff is instantaneous, and the main use of water is for energy. In the west, runoff is slow, and the major use of rainwater is irrigation.

Working together as a team, the community can advance the understanding of water-cycle extremes. The next steps are to (1) improve cross-modeling interoperability and build the water cycle into the CESM components, most notably the Community Land Model; (2) couple with the Community Land Model, Global Change Assessment Model, and ocean and ice components. Topics will selectively be included in future solicitations by DOE and other agencies. Interagency working groups that will advance the workshop's results include the Interagency Group on Integrative Modeling of the U.S. Global Change Research Program (USGCRP) and the National Climate Assessment.

Baldocchi said that he did not see any mention of Ameriflux in the report. The hydrology program should make use of Ameriflux's outputs. Joseph said that it was discussed at the workshop but must have been missed in the report.

Stacey noted that, in the federal government, NOAA has responsibility for water policy. He asked if NOAA was connected to this workshop. Geernaert stated that no one agency dominates water policy; NOAA, the U.S. Geological Survey, and the U.S. Army Corps of Engineers all have interests in water policy. The MOU with NOAA did not get into great detail on topics, but water is a hot topic, and our understanding is weak on this issue.

Washington pointed out that the National Climate Assessment being published by the USGCRP will be very helpful to government agencies struggling with climate change. However, climate change is only one of the stressors that have to be dealt with. The report does address water policy and use changes. This will be an area of top interest.

Ehleringer noted that humans' social systems play a large role in water use. He asked if this effort were an extension of DOE's interest in human social systems. Joseph replied that models need to be advanced further to understand water systems. Geernaert asserted that there is no philosophical change on DOE's part. In the past, the physics system has been focused on. But one cannot be limited to just physics systems; one needs to understand human systems, such as dam building, land clearing, and human settlements. Ehleringer asked where models will be in a decade and what data will they need at what resolution. Geernaert replied that he did not know what research will be needed but did recognize that all the bases, including human systems, need to be covered.

Remington noted that similar discussions had been held when the National Ecological Observatory Network (NEON) was started. NEON's infrastructure should be more integrated with the modeling efforts. Joseph replied that the participants list was confined to those doing modeling of hydrologic effects. Weatherwax added that the Office recognizes that NEON has similar interests and has been working with the personnel there.

Wanda Ferrell was introduced to report on the Atmospheric Radiation Measurement Program/European Union (ARM/EU) Workshop.

The workshop was held November 6-8, 2012, in Washington, D.C., with 36 scientists invited, half from the United States and half from Europe. Observational, process-research, and modeling experts from eight countries were represented. The workshop started with four formal talks and the statement of the science questions that were to be the basis of discussion for the following two days:

- What is the distribution of aerosol properties for the Atmospheric Model Intercomparison Project (AMIP) period (i.e., since 1979)?
- What is the coupling among microphysics, aerosols, and cloud dynamics as a function of scale and regime (e.g., vertical velocity or stability)?
- How are precipitation, water vapor, and cloudiness coupled, and what roles does organization play in this coupling?
- How do clouds and precipitation couple with surface properties?
- What is the response of clouds to warming?
- What is the response of the probability density function of precipitation to warming?

The workshop was designed to address the geophysical variables needed, including accuracy and resolution needs; the type of correlated data sets needed to address the scientific questions; and the best mix of laboratory, campaign mode, and long-term data sets.

The key workshop recommendations were to (1) establish a bilateral steering committee to develop collaboration mechanisms and instruments, oversee parallel bilateral working groups, and develop and execute strategies of common interest; (2) establish a set of six working groups (WGs) to coordinate among the key DOE and EU ground-based remote-sensing centers on radar calibration, microwave radiometry, retrievals, integrated data portals, initialization data sets, and operational use of large-eddy simulations; and (3) coordinate participation in major field campaigns (i.e., GOAmazon, Atlantic Observations, Arctic Sea Ice Study, and Southern Ocean Observations).

Since the workshop, a steering committee has been established to ensure progress and to coordinate bilateral team meetings, develop strategies, and execute plans. It established the six working groups, each of which has U.S. and European cochairs. The working group on radar calibration has the goal of combining radar experts from the ARM and European observatories to initiate common traceable methods for cloud-radar calibration. The working group on microwave radiometers has the goal of developing a set of collaboration mechanisms that leads to a common procedure on the data flow of operational microwave-radiometer measurements. The working group on retrievals has the goal of combining data-retrieval experts from the ARM and the Cloudenet, the European Ground-Based Observations of Essential Variables for Climate and Operational Meteorology (EG-CLIMET), and High-Definition Clouds and Precipitation for Climate Prediction [HD(CP)2] communities and developing mechanisms to enhance collaboration on the retrieval of the cloudy, thermodynamic atmospheric state through ground-based remote sensing. The working group on large-eddy simulation (LES) has the goal of combining ARM/ASR and European expertise and evaluating the operational use of LES at supersites and during field experiments. The working group on integrated data portals has the goal of combining experts involved in both ARM and EU facility data informatics; developing the architecture, standards, and framework for an integrated portal; and documenting the metadata, products, and related information. That working group has drawn up a portal-architecture proposal for which a number of technical and policy issues will have to be resolved. The Europeans needed a metadata editor to input their data into the integrated system; an editor developed in Oak Ridge will be used. The working group on initialization data sets has the goal of combining researchers from both the global climate modeling (GCM) and LES communities, determining which specific data sets from ARM and European facilities can be used to improve initialization; and evaluating process and prediction models. The working group on workshops will convene workshops by PIs to discuss how to proceed on each campaign.

Penner asked how the LES model would be used in an operational sense. Zhang said that the data are there, and the models can be used in a predictive capacity.

A break was declared at 9:24 a.m. The meeting was called back to order at 9:44 a.m.

Adam Arkin was asked to present an update on the DOE Systems Biology Knowledgebase.

In 2010, the need for a knowledgebase in biology (KBase) was identified, one that would reflect and support investigative tools and research efforts. DOE was the natural home for such a system because its mission is to predict, control, and design biological components of energetic processes and environmental balance. These are very complex processes, and the science has become very complicated. A major goal is to understand how microbial communities interact with the environments affected by energy production and use.

The community set specifications for KBase in 2011: a federated infrastructure and 10-Gb/s data transfer capabilities; 2 PB of data storage; 2000 cores for data processing; an interface between high-performance computing and cloud computational resources; and management tools including application, semantic, and user interfaces. For microbial systems, the system was to be able to reconstruct and predict metabolic and gene-expression regulatory networks to manipulate microbial function. For plant systems, the system was to be able to integrate phenotypic and experimental data and metadata to predict biomass properties from genotype and to assemble regulatory data to enable analysis, cross-comparisons, and modeling. For microbial communities, the system was to be able to model metabolic processes and mine metagenomic data to identify unknown genes.

Today, KBase has a federated infrastructure, a 10-Gb/s transfer capability, 12,000 cores for data processing, and more than 3 PB of storage. The integrated KBase is application-programming-interface (API) specified and operational. The integrated data model is aware of 925 data types. Some 40 interface description documents lead to 821 functions that can be compiled into use for PERL, Python, Java, and R. There are Prototype Search, Workflow, and Novel Narrative/Notebook interfaces for navigating, analyzing, and building knowledge in KBase.

For microbial systems, it has metabolic and regulatory reconstructions for 5,534 prokaryotic and 161 archaeal genomes; 7,830 genome annotations with 23,058,670 features predicted; 12,620 regulons with 266,345 protein families inferred; 4,985 metabolic models; 6,202 growth curves, 1,947,690 strain-fitness measurements, and 3,227 gene-expression data sets; and services for assembly, annotation, phylogenomics, regulatory and metabolic network inference, flux-balance analysis (FBA) and probabilistic regulation of metabolism (PROM) modeling of metabolism, and reconciliation and improvement of models against data.

For plant systems, it has more than 175 eukaryotic genomes; phenotypes for genome variants of plants; services for calling the genetic variation among individuals; services for variation calling; mapping of genotype-to-phenotype via genome-wide association studies (GWAS) analyses; tools for candidate-gene filtering, trait modeling, and pathway enrichment; 731 gene-expression experiments; plant co-expression network analysis; and initial plant metabolic modeling.

For microbial communities, it has access to 11,000 metagenomes; access to QIIME (Quantitative Insights Into Microbial Ecology) Pipeline functionality; new tools for metagenome-sequence quality assessment and experimental design; and services for taxonomic and gene identification, abundance, and a host of other functions.

The goals are to solve the grand challenges in biological control of environmental and energy processes, to deal with deep issues in scientific communication and reproducibility, to lower the barriers to computationally efficient use of advanced algorithms and data from diverse producers, and to do this by leveraging many minds.

KBase's bioinformatics resources include an increasing number of data warehouses, specialized applications and databases for relatively generic analyses, evolving libraries of sophisticated computational biology algorithms for use in programming environments, workflow tools that allow the chaining of these algorithms by nonprogrammers, workflow-sharing tools to allow people to use each other's work products, open-access publication of journal articles with increasing use of semantic tagging, and scientific social networks.

KBase drives data through models to predictions and experimental design; accelerates reproducible, reusable, and transparent science; deeply enables scientists to work together to approach complex biological problems; and gives credit where it is due and privacy where it is needed. It is an open software and data environment to which others can contribute and with which others can build.

Users can access data and tools through the command line, API, and browser to produce integrated data for predictive models, hypotheses, visualizations, and comparative analyses to conduct analysis, simulation, and prediction. With KBase, one can now transparently access multiple heterogeneous datasets and bioinformatics tools; efficiently annotate new microbial genomes; infer metabolic and regulatory networks; transform network inferences into metabolic models; map missing reactions to genes using novel data-reconciliation tools; design effective sequencing strategies for complex multi-sample metagenomic projects; test microbial ecological hypotheses through taxonomic and functional analysis of quality-assessed metagenomic data; predict plant gene function and molecular phenotype; discover genetic variations and map them to complex organismal traits; and share data, analysis tools, workflows, and scientific conclusions with the community.

The overall goals are to reconstruct and predict metabolic and gene-expression regulatory networks to manipulate microbial function, vastly increase the capability of the scientific community to communicate and utilize their existing data, and enable the planning of effective experiments and maximize the understanding of microbial-system function. KBase allows one to rapidly build an understanding of microbe behavior from the available genome sequence assembly data and annotations by doing metabolic model reconstruction, phenotype simulation, and phenotype reconciliation. KBase is driving discovery and experimental biology by integrating all of these activities and validation experiments and then reiterating the process.

KBase supports massive metagenomics. The targeted genomes are poplar, arabidopsis, sorghum, chlamydomonas, brachypodium, miscanthus, and switchgrass. It vastly increases the rate at which people can conduct an analysis. It allows networked-based knowledge discovery, facilitates building predictive models, and provides infrastructure for scientific social networks.

The KBase framework was used to enable PROM, which integrates metabolic roles and functional data to make better predictions of growth given genetic or transcriptional variation. KBase has been used to identify important single-nucleotide polymorphisms (SNPs). The KBase workflow model drives data toward dynamic models of function so the user can build a research path and move through it in a reasonable way. It saves the researchers' sessions and shares the results, which can be accessed on the system or downloaded to a desktop. The KBase labs are prototypes and applications that demonstrate KBase's present and possible future functions. KBase's IRIS [interactive remote invocation service] command line can be used on the web to provide a command-line environment for programmers. It is also downloadable for desktop use.

Prototype collaborative workspaces allow one to view objects produced by analyses. They may be used for all public and private models and FBA solutions. The search prototype does keyword searches of the literature, genomes, function, or regulation.

There is a large range of initial platform modules, incorporating graphic user interface (GUI) components, services, infrastructure services, data stores, language support, and cloud virtualization. Next year, there will be hundreds more. Already, there is a wide range of KBase services.

KBase stores a diverse range of biological data in the form of highly structured data in relational databases, frequently changing user data, and large bulk data, providing petabytes of raw data, flexible storage for workspaces, and structured storage for curated data. Assembly services include the control server, workers, job queues, authentication, and data repository.

Several types of training are being offered, such as user manuals, tutorials, workshops, and boot camps. Online resources also include video tutorials, developer documentation, a calendar of events, frequently asked questions, and press releases.

KBase has coordination and shared milestones with the BRCs; infrastructure collaborations with the Joint Genome Institute, Environmental Molecular Sciences Laboratory, and National Energy Research Scientific Computing Center; scientific collaborations with ENIGMA [Enhancing Neuro Imaging Genetics Through Meta-Analysis], Plant Microbes Interface Project, and other SFAs; and MOAs with NSF for iPlant and with the USDA (pending).

In the future, the data model must evolve to support the modeling mission; a framework for turning bioinformatics algorithms' output into models needs further development; a theory for integrative, cross-scale predictive biology is under development; much better ontologies are needed for nearly everything; the social tools must be built; data import, quality assessment, and metadata need to be improved; new third-party algorithms and support with scalable computing need to be incorporated; a strong external development community must be grown; and the KBase Foundation must be launched to ease licensing and growth of KBase user participation. The concept is to develop interactive community knowledge through the ability to automatically record workflows and findings so other researchers can easily review this information and build on it.

Schlicher asked two questions: (1) What might be done with the USDA, and (2) can one see any hope for agricultural supply and demand models, which do not include environment or plant components and therefore limit predictive capability? Arkin replied that the USDA has a few programs that are really important, a beautiful quantitative breeding program, and a lot of animal models of microbial communities that affect the animals that DOE would like to have access to and be part of. On the climate side, it is known that microbes and plants change the global nutrient cycle and affect such processes as erosion and light harvesting. When there are measurable metrics, KBase can support users; the field is not at the point where KBase can easily expand to deal with the size of things that they are dealing with.

Curry asked how much up-scaling would be needed for the results to be meaningful to the ARM Program. Arkin replied that there is a question whether the computer infrastructure can support more users and algorithms; it may be stressed if additional teams are added. There are two leaderships (Kbase and ARM), and they may have different visions, and the coordination of the two groups may not scale. This problem might be solved by the use of specifications and a closely knit partnership between Kbase and the partner program (e.g., ARM). Curry asked about the feasibility of the scope. Arkin replied that it has to be done incrementally. One could get stuff going quickly, perhaps in 5 years to scale up to climate change.

Baldocchi called attention to the Toxics Release Inventory (TRI) database which, he said, would be a rich resource for KBase.

Remington asked how the KBase fitted into the virtual laboratory, noting that having something that helps non-computer-scientists to use the system is important. Arkin responded that one of the great deliverables is the ability to build an application that others can then use without any programming needed. Remington commented that data integration is not going to make things happen magically; this is not a data-integration center but a collaboration center.

Stahl said that the narrative produced in the course of using KBase is in the human dimension. He asked if there were going to be any way to set up a privacy wall. Arkin answered, yes; researchers can be given tools much like those of social networks. Moreover, whatever a user decides not to tell KBase does not appear in KBase. The project is still thinking about things like this. It is desirable to enhance sharing, though. One might start considering the narrative a publication.

Wall conjectured that KBase could replace laboratory notebooks. Arkin replied that he hoped so. Even if the system went down, the information would be retained.

Joachimiak asked how these hypotheses got validated and about sector data. Arkin responded that one would have to come back and add the validation narrative. Unvalidated hypotheses would be exposed in the KBase. Pathway design needs sector data. Sector data needs to be incorporated.

Washington asked how they would know that no mischief was going on. Stevens answered that there is a security enclave across the four involved national laboratories and ESnet. Open authorization is used, and it is very solid. There is a cybersecurity plan that has been adopted by all four national laboratories. All primary servers are replicated in three other places, the network is very fast, and all this is transparent to the user.

Ehleringer asked if the system can be overwhelmed by the size of the user population. Arkin replied that there are now about 20 collaborators using it. That number will increase by a couple of hundred per month and then increase rapidly to about 70,000 users. Stevens said that the system has been tested with a couple of hundred users. Arkin admitted that there will be slowdowns as the user base increases and the system adapts to that level of users.

Merchant asked what was happening with data acquisition. Arkin responded that there are staffers to acquire data, and users can upload data themselves. There are boot camps where users can learn how to do this.

Stacey opened a continued discussion of the facilities charge. A straw-man response has been drawn up.

Hubbard liked the systems-biology approach. The two-page letter should look like the last BERAC report's Executive Summary. The case needs to be made up front. The integrated framework should be highlighted.

Baldocchi asked what the topics were of the systems science networks and predictive science. He suggested "biosphere," perhaps.

Remington said that the paragraph that had been added did a good job of clarifying things. The National Center for Biotechnology Information (NCBI) has a $250 million budget; there is not much funding available from them to sustain this effort.

Shaver suggested that something about implementation be added at the end. Some components are available now and need to be improved, and some need to be created.

Schlicher complained that the magnitude of the opportunity available with the $100 million investment does not come across. Stacey pointed out that BERAC is being asked to approve something more than what DOE originally proposed, including a systems-engineering component. Remington said that the Subcommittee considered the whole system, not just the CASK component, which was considered ripe for early adoption.

Penner noted that how the goals are stated could be firmed up; some specific language would be e-mailed to the chair.

Ehleringer was not impressed with the language. "Climate change" should be linked in.

Hubbard asked if the letter had to address readiness. Stacey agreed that the charge called for a ranking and assessment of readiness. The community is stepping toward this goal with KBase. Hubbard asked what "shovel ready" meant. Weatherwax said that it is known what infrastructure would be needed and that the cost would be over the threshold for a major facility. Remington commented that the call for workshops to decide what to do belies the fact that this is "shovel ready." Stacey explained that the workshops are for community buy-in. The language may need to be tweaked. Penner said that stating that there is a pilot project (KBase) would indicate that the community is ready for some of this.

Joachimiak said that it should be stated that this will allow people to do things that they currently cannot do. The big picture should not be lost sight of. Shanks suggested that one could say that one node already exists.

Shanks said that the implementation details and a description of the ultimate structure would lend more confidence to the proposal and underscore that these things *are* possible. He wondered how many people are going to use this and whether the structure was scalable. Remington pointed out that KBase is focused on genomic data, which is very meager, but a lot of other components will be added that are more complex. The ability to deal with this non-homogeneity is important to relate.

Stacey suggested forming a letter-writing sub-subcommittee to produce the final draft in the next few weeks. That draft could be circulated by e-mail and tweaked and voted on during the 1:00 p.m. EST, March 7, teleconference. The consensus of the Committee was to adopt this procedure.

The floor was opened to public comment.

Janet Jenson of LBNL stated that one should define one's goals in terms of science drivers. To solve this grand challenge will require an infrastructure. An example would be a science driver for, at a global level, identifying the microbial contributions to the production of greenhouse gases (both as sources and sinks). Perhaps one could describe three of four science drivers first and then describe infrastructure. Stacey said that the Subcommittee assumed those science drivers in its discussions.

There being no further public comment, the meeting was adjourned at 11:30 a.m.