

National Science Foundation
Directorate for Biological Sciences

Advisory Committee Meeting
April 22-23, 2015
Hilton Hotel, NSF Room 1235

Summary Minutes

BIO AC Members in Attendance:

Hannah Carey (4/22)	Wilfredo Colon	May Berenbaum
Greg Florant	Steve Goldstein	Katherine Gross, Chair
Linda Hyman	Elizabeth Kellogg	Susan Marqusee
Richard McCombie (4/22)	Margaret McFall-Ngai	Randy Nelson
Michael Purugganan	David Schimel	Joan Strassmann
Brett Tyler		

BIO AC Members not in attendance: Drs. Gaetano Montelione, Wendy Raymond, Stacia Sower, and Paul Turner

Wednesday, April 22

Welcome

Dr. Katherine Gross opened the meeting by welcoming everyone and asking for introductions of the committee members, NSF staff and guests.

The summary minutes from the September 2014 BIO AC meeting were approved.

Dr. Gross reviewed the meeting agenda.

Budget Review (FY15) and Budget Request (FY16) – Dr. James Olds, Assistant Director, Directorate for Biological Sciences

Dr. Olds summarized the NSF FY15 Budget and the BIO FY16 Budget Request and discussed the breakdown of the request across the directorate and the Foundation. Core investments remained focused around the Five Grand Challenges for 21st Century Biology. These include Understanding the Brain (UtB), Synthetic Biology, the National Ecological Observatory Network (NEON), Research at the Intersection of Biology, Math & Physical Sciences and Engineering (BioMaPS), Innovations at the Nexus of Food, Energy and Water (INFEWS), and Cyberinfrastructure for the 21st century (CIF21). Dr. Olds also described ongoing BIO Data activities and possibly developing plans for a National Brain Observatory.

The Advisory Committee discussed the National Brain Observatory, the funding trajectory, partnering with other agencies, synthetic biology, and the position of BIO in the funding profile of NSF.

Updates and Emerging Issues:

INFEWS – Alan Tessier

Dr. Tessier began by stating that this new initiative could benefit from input from the BIO AC. He described the evolution of INFEWS from the Science, Engineering, and Education for Sustainability (SEES) program – a multi-year NSF-wide activity that grew dramatically into a complex set of programs. In this context, the importance of understanding the interconnected and interdependent systems involving food, energy, and water became apparent and two new activities emerged: INFEWS and Risk and Resilience, which together amount to more than \$110M. Risk and Resilience programs focus on technology and are supported primarily by CISE, ENG, GEO and SBE. INFEWS is focused on addressing the societal challenges of food production for a growing population, as well as increasing demands for energy and water. Dr. Tessier reported that the NSF-wide Advisory Committee for Environmental Research and Education (AC-ERE) has recommended that NSF maintain a stable, long-term investment in environmental research by building upon existing programs and beginning new initiatives. He went on to point out that although sustainability has become ingrained in most NSF programs, NSF is confronting challenges of workload and leveraging investments in infrastructure.

The BIO AC discussed: the potential for including aquaculture and near shore environment research in INFEWS; overlap of Risk and Resilience groups and INFEWS; interest in how insurance companies deal with risk and resilience; exclusion in INFEWS of the interactions with the natural environment; future of the programs included in the SEES initiative; biotic extreme events; and climate change.

Dr. Gross asked the BIO AC to consider what would be the most effective ways for engaging BIO in the development of INFEWS.

Overview of INCLUDES – Dr. James Olds

The Inclusion across the Nation of Communities of Learners that have been Underrepresented for Diversity in Engineering and Science (INCLUDES) initiative is an NSF-wide initiative that will address Broadening Participation in Science and Engineering (S&E). Dr. Olds discussed prior investments by NSF in this area and the uncertainty surrounding their effectiveness, including methods to measure the outcomes. Dr. Cordova has required that INCLUDES be NSF-wide, coherent and scalable. Dr. Olds posed the questions: How should NSF go forward? How can NSF do better?

The BIO AC engaged in a broad discussion of broadening participation in S&E with particular attention to the impact of NSF investment in broader participation activities, especially any that could scale up to general target groups; data driven analyses of hiring practices and retention in academia, as well as professional re-entry after a necessary career interruption; and career development outside of academia. :

Transparency and Accountability – Dr. Jane Silverthorne

Dr. Silverthorne presented the transparency and accountability (T&A) talking points from Dr. France Cordova's presentation to the National Science Board. Congress has requested NSF provide evidence of transparency and accountability in the merit review process. In response, NSF has

- released important notices (No. 135, 136, and 137);
- changed instructions related to abstracts;
- launched a website dedicated to T&A;
- established a working group; and
- created a liaison in the Office of the Director.

Dr. Silverthorne described the activities being initiated in BIO to develop and implement greater transparency and accountability in the merit review process, including establishment of a BIO Portfolio Analysis Working Group (JS is the liaison) and development and implementation of a template for Division contributions to a BIO-wide Annual Report.

The AC discussed a number of activities that could improve the transparency to the public of the research supported by BIO and its importance. Several of these could be implemented NSF-wide including a public facing transparency website, training for PIs to communicate with the public, and a science marketing program to improve public trust in science. The important role of the AC in communicating to peers and professional societies was also discussed.

Reproducibility – Dr. Charles Liarakos

Dr. Liarakos summarized recent activities related to reproducibility and replicability, including recent workshops, a data science track of the new NSF Research Traineeship (NRT) program for graduate students, and the response by all NSF Directorates to questions about reproducibility in their discipline from the Office of the Director. Dr. Liarakos described recent meetings that focused on data reproducibility and repeatability: a December 2014 meeting in UK, chaired by Dr. Francis Collins, NIH; a Wellcome Trust meeting in April 2015; and a meeting with the House Science Committee staff at NIH (BIO, SBE, and EHR). Dr. Liarakos framed the general discussion by noting that concern about reproducibility of research results began with clinical and behavioral studies, but that in biology, reproducibility does not and cannot imply exact replication of results.

The BIO AC engaged in an extensive discussion with particular attention to general variability of reproducibility in science and the lack of exact reproducibility as a consequence of the evolution of knowledge; standards and good practices for reproducing research; and the need for accurate and complete publication of “methods” in journal articles.

There was a brief discussion between BIO AC members and BIO leadership about the role of the BIO AC and a reconsideration of the format of the meetings.

Graduate Education – Introduction of NSF’s *Investments in Graduate Education: Five Year Strategic Plan*; formation of AC sub-committee for follow-up? – Dr. Charles Liarakos

Dr. Liarakos summarized NSF investments in graduate education and support. NSF supports approximately 40,000 graduated students by three primary mechanisms: research awards (85%), fellowships (10%) and training grants (5%). He described a blog site, NSF Forum on Graduate Education (<https://nsfgradforum.wordpress.com>), created by the Education Directorate, where graduate students could share their experiences, and noted that an NSF-wide committee was currently developing a strategic plan for the NSF’s support of graduate education. Dr. Liarakos presented two questions for the BIO AC follow-up discussion:

- 1) What is the ‘right mix’ of sources of support to graduate students?
- 2) Can NSF structure its support of research awards to better track the return on investment?

The AC members discussed funding rates, the need to expand master’s degree support, the academic culture regarding graduate student training, tuition costs, training for a wider range of careers, and institutional requirement versus PI burden.

Dr. Gross suggested that the BIO AC discuss the various options with BIO staff at the next AC meeting.

Public Access – Dr. Jane Silverthorne

Dr. Silverthorne presented the requirement by OSTP that all agencies should have plans for increasing access to results of federally funded scientific research. NSF's plan, which includes software access, was delivered to OSTP on time and accepted in 2015. In response to input from broad stakeholder groups, NSF decided to use the PAGES system, an existing DOE system for tracking publications that is compatible with NSF systems. Dr. Silverthorne described the internal process, which was developed with input from the NSF Steering Committee, technical teams from across NSF, and Program Officer working groups. The guiding principles of the NSF plan are to provide public access, minimize the workload burden on awardees/investigators and NSF staff, embeds public access activities in existing workflows, leverage existing systems, be extensible to other products of NSF-funded research, and minimize cost. Dr. Silverthorne further described what will and will not change for PIs, and the next steps in the implementation. An article about the plan has been posted on BIO Buzz, BIO's blog, as an initial step in communicating with the research community. Dr. Silverthorne asked the BIO AC members to try out the beta repository and encourage others to do so. She concluded by describing the FY14 pilot awards funded by BIO to support public access to data.

The BIO AC raised some concerns regarding this report including the 12 month embargo waiver, capturing metadata, connection of data repositories with search mechanisms, using a DOI to automatically fill in information, juried conferences, and comparison of the NSF and NIH processes. A progress report on BIO compliance with the NSF Public Access plan will be provided at the next meeting.

BioData: setting the stage for meeting with ACCI – Dr. James Olds

Dr. Olds discussed the increasing number of large data sets in biology and the need for greater processing power. He asked the BIO AC for suggestions as to how NSF should deploy its limited resources to best handle these data and computational demands in a way that leverages previous investments but still drives the science forward. Currently, NSF-wide investments in computing and data management are not well coordinated and need to more efficient. How can NSF manage the evolving data and computational demands of biology?

The BIO AC agreed to focus on two questions for the joint session with the ACCI:

1. How the NSF should think about investments in computational and data management that would build upon what has been done?
2. What should the framework be for handling computational and data management needs that aligns with NSF policies and interests?

Preparation for meeting with NSF Director, Dr. France Córdoba

A number of topics were discussed for the meeting with Dr. Cordova and the consensus was to focus on her vision for broadening participation, the role of BIO in the basic research enterprise of the nation (especially in relation to the NIH), the state of American research competitiveness and the morale of the American scientific community, and communicating the effect of NSF-supported transformational basic research to the public and the Congress.

Meeting with Dr. Córdoba

Dr. Olds introduced Dr. Cordova and Dr. Gross welcomed her and asked the BIO AC members to introduce themselves.

Dr. Cordova thanked the BIO AC and expressed pleasure that Dr. Olds had joined NSF as the BIO Assistant Director. She also discussed NSF participation in the Congressional hearing on the BRAIN

initiative; the recent AD retreat; her vision for BIO, developing the best possible relationship with both the Senate and House, and trying to engage all parties by providing necessary information.

Dr. Cordova and the Advisory Committee discussed: interagency collaborations, such as Understanding the Brain and INFEWS; the mid-level tenure process; the funding success rate of women and under-represented groups; NSF's overall funding rate; tension between Congress and NSF based on societal concerns versus science goals; the contribution of NSF and NIH to the nation's competitiveness; BIO AC members' roles in effectively communicating the effect of NSF-supported transformational basic research; NSF-wide efforts in Broadening Participation; INCLUDES; improvement of science communication; PIs sharing their successful research outcomes with Congress; and concern over NSF's underutilization of the website as an effective tool for communication.

Dr. Gross concluded the session by thanking Dr. Cordova for her time and the efforts she is making to communicate the work from BIO and NSF to Congress. The BIO AC is happy to help with these efforts.

The meeting was adjourned for the day.

Thursday, April 23

The BIO AC meeting was re-convened at 8:30 AM by the Chair.

Joint session with Advisory Committee for Cyberinfrastructure (ACCI)

Dr. Peter Arzberger began the session by noting the passing of Dr. John C. Wooley, and recognizing Dr. Wooley's long commitment to NSF. Dr. Wooley was the founding Division Director of DBI and served on the BIO AC. He provided a brief summary of Dr. Wooley's career and contributions to science and the NSF

Dr. Olds welcomed the members of both ACs. He discussed the synergies between the biological and computer sciences and their unique potential to advance science. He encouraged conversations between the ACs as a way to identify new opportunities.

Dr. Irene Qualters, DD of the Division of Advanced Cyberinfrastructure (ACI) provided an overview of the scope of ACI and its mission: to support advanced CI to accelerate discovery and innovation across all disciplines. She stated ACI supports use-inspired research and applied research, has a coordination role across NSF, and is inherently multi-disciplinary with strong ties to all divisions and directorates.

Discussion topics:

What would it take for iPlant to "talk" to NEON and what might that mean for the science?

Overview - Drs. Karen Cone and Elizabeth Blood

Drs. Cone and Blood provided overviews of the capabilities of the iPlant Collaborative and the National Ecological Observatory Network (NEON), respectively. Developed with a focus on addressing grand challenges in biology, iPlant is a virtual center based at the University of Arizona that integrates high power computing resources to support data access/research projects in Life Sciences, provides data storage, and supports data analysis. NEON is the first continental scale research observatory designed to advance fundamental theories of life. NEON combines data collected through biological sampling and analysis, sensors, and sensor networks through a centralized cyberinfrastructure backbone to create a nationwide network. The presentation also focused on the added value to the wider research community at the intersection of iPlant and NEON, which would focus on reuse and interoperability to leverage both platforms effectively.

The discussion that followed covered a wide range of topics including: sustainability plans for iPlant; common data formats and interoperability; leveraging other supercomputing centers; the ability of both platform cyberinfrastructures to support reuse of data and quick adaptation; coding NEON data to work on other infrastructures (make data available and searchable); the amount of NEON data; iPlant use by the research community; and the relationship of the two platforms to DataOne.

IDigBio: Aggregating, re-formatting, and federating many small collections? Where is the science?

Overview of iDigBio – Dr. Anne Maglia

Dr. Maglia described the development of the Advancing Digitization of Biological Collections (ADBC) initiative that led to iDigBio. The goal was to coordinate and make available to users the information that is stored/recorded in biological collections. Approximately one billion biological specimens have been collected of which 27.6 M specimen records have been digitized through the iDigBio program. iDigBio serves as the hub for the thematic research networks and captures the data based on the themes funded through the ADBC program. ADBC currently funds digitization at 202 institutions across the U.S. and Guam and includes Canadian partners. iDigBio has 6 primary activities: 1) implementation of a cloud-based data store, 2) community building, 3) the development of tools, APIs & workflows, 4) outreach/education 5) developing data capture, annotation, and best practices, and 6) research use/training. Dr. Maglia described the cyberinfrastructure design of iDigBio and how from inception iDigBio was built to “integrate, leverage, and reuse” data. Recognition of biological collections as a national resource and leveraging existing specimen records in collections is vital. Dr. Maglia presented examples of the analysis of iDigBio data that have led to discoveries from recently funded projects. She discussed future requirements beyond what iDigBio and ADBC currently make possible and bridging investments such as ARBOR, Open TOL, Lifemapper, Phenomics, and BIEN.

The extensive discussion that followed included topics such as: developing a system for tracking changes in specimen records; use of data to validate models, particularly niche models; “iDigBRAIN” (centralized storage of digitized images of the brain); inclusion of international specimens; integration with GBIF; inclusion of crop genomics data; concern that NEON is not thinking about data integration and reuse from inception; scalability of systems; incentivizing collaboration versus incentivizing success; a mechanism for receiving credit for use of a PI’s image/data; the benefit of a collection’s data being used; probability of a specimen getting into a collection; sampling bias; use of iMicrobe data; alignment with the Critical Zone Observatory data systems; parallels in the earth science community; definition of “reproducibility”; different types of reproducibility (computational, empirical, and statistical); consistent standards for embedding citations and incentivizing to use the standards; transparency of the data chain of evidence and data validation; and workshops on reproducibility.

Wrap-Up – Drs. Katherine Gross, BIO AC Chair and James Olds, BIO AD

Potential dates for the next AC were discussed: September 9-10, 12-15, or 27-29, 2015. A poll will be taken to find the date that will allow most AC members to attend. The need for setting AC meeting in advance was also discussed.

Dr. Gross thanked the AC members for participating and noted the recent award of the Medal of Science to May Berenbaum and the election of Margaret McFall-Ngai to the National Academy of Science.

Dr. Gross asked for input on what issues and activities should be the focus on the BIO AC and can be the focus of future meetings (and work of sub-committees). The AC agreed on the following:

- Graduate Education and Training;
- Broader Impacts;
- Broadening Participation (and its relationship to BI);

- Data Accountability and Reproducibility; and
- Strategic Plan development.

Dr. Olds discussed his plan to present his vision for the Biology Directorate at the September BIO AC meeting and the development of that plan, including engagement with the BIO AC members, scientific community stakeholders and other partners.

The Chair adjourned the meeting at 2:00 pm.