Response to Recommendations
by the Emerging Frontiers Committee of Visitors
February 21, 2014

The Directorate for Biological Sciences (BIO) expresses its sincere appreciation to the members of the Emerging Frontiers Committee of Visitors (COV) for their incisive observations and constructive recommendations. It was evident that all the members of the COV were deeply interested in the welfare and development of the programs in Emerging Frontiers and the science communities that are served by these programmatic activities.

In general, the COV report presents a favorable evaluation of the operations and management of the MacroSystems Biology (MSB) and Advancing Digitization of Biodiversity Collections (ADBC) programs in Emerging Frontiers (EF), though a number of important recommendations were given to improve current activities and practices.

BIO is particularly appreciative of the COV's recognition of the EF staff's conscientious efforts to handle the review process, from the selection of the reviewers to the funding decisions, with great care and integrity. Furthermore, BIO also appreciates that the COV found that both programs have sustained balanced portfolios across different topics and biological systems and have included "risky" projects in their award portfolios. The programs will strive to realize emerging opportunities through continued collaborative efforts across all levels of BIO and with the rest of NSF.

Finally, BIO recognizes as important and timely the COV's emphases on the necessity to focus on data infrastructure, interagency collaborations relating to data and collections, and development of tools for utilizing the data.

SUMMARY RECOMMENDATIONS

Recommendation 1 (page 2): NSF should take a long-term strategic look at its data infrastructure needs and plans. For each of these programs, access to data external to a project is of substantial scientific importance and allows the science to advance appreciably. Especially important is to find a way to assess the status of the multiple data efforts that are occurring across the agency (e.g., DataOne, NEON, iPlant, etc.). Perhaps a partnership with EarthCube could develop interoperability, community governance, and other capabilities that are general to the environmental sciences.

Response: BIO is assessing the data integration of many programs and has tasked several staff members within the Directorate to examine data-relevant issues, needs, and potential solutions. Realistically, "big data" is beyond the resources on any single program or Directorate and therefore is being addressed agency-wide. It should be noted that ADBC has established an agreement with the U.S. Geological Survey (USGS) to incorporate all of the data products from awards made by ADBC into Biodiversity Information Serving Our Nation (BISON), the U.S. federal resource for biological occurrence and collections data. Additionally, the ADBC program solicitation notes that any project involving specimens of importance to the Geosciences (GEO) should include a description of that importance to EarthCube. Moreover, EarthCube funds have supported some of the ADBC projects to assure interoperability of the data with GEO efforts. The issue of data and data infrastructure also has been a recurring topic within the MSB program and has been addressed in the Frontiers in Ecology and the Environment Special Issue from February 2014. Both programs will strive to ensure the needs of their particular communities are articulated in the larger efforts.

Recommendation 2 (page 2): The COV recommends that NSF continue to pursue interagency collaborations relating to collections. Given that iDigBio is now operating and many collections are being funded to digitize and deposit their data in iDigBio, it seems time for an interagency approach, including digitization of federal collections and the expansion of the central database.

Response: BIO agrees with this recommendation. As noted in the response to Recommendation 1, all data from ADBC awards will be integrated into BISON. Most biodiversity collections (including all currently supported by ADBC) also feed their data to the Global Biodiversity Information Facility (GBIF) directly; for those collections that do not, iDigBio provides the data to both GBIF and BISON (BISON operates the North American GBIF node.) Finally, since the primary focus of the Interagency Working Group on Biological Collections has been on issues pertinent to standards and policies for federal collections, issues related to data have yet to be thoroughly discussed. This interagency working group has designated BISON as the host for all federal collections data and has included the head of BISON as a participating member. A separate interagency working group (led by BISON), assessing data related to biodiversity and the environment, is considering larger issues of data integration and provides updates to the ADBC Program Directors. Upon receipt of this recommendation from the COV, three members of the NSF ADBC working group met with the director of the Smithsonian Institution's National Museum of Natural History. They began discussions for setting standards for the interoperability of the Smithsonian's collections with the digitized data resulting from ADBC projects that will continue for the next few months.

Furthermore, the ADBC program always has included at least two federal agencies on its annual review panel so that discussions about the new workflows being developed in the community can be transmitted to other agencies. This year at least two jointly sponsored workshops have been held with iDigBio and the Smithsonian Institution or USGS. These workshops have helped to provide integration ideas and innovative approaches to the digitization of collections.

Recommendation 3 (page 2): The COV believes that it is time for NSF to initiate a plan to develop the tools that will be required for utilizing the massive specimen and environmental databases to their full potential. Just as the sequencing of the human genome demanded new analytical, computational, and sequencing tools, so too will the growing database require creative efforts by computer specialists, engineers, and a variety of scientists to determine how these data can most efficiently and effectively be utilized across disciplines.

Response: BIO acknowledges the importance of this issue but also recognizes that it is beyond the scope of one directorate. BIO is aware of the need for new ways to analyze increasingly large data sets. BIO is currently supporting the development of tools for data synthesis and analysis through its centers and through the soon to be operation National Ecological Observatory Network (NEON). Because the focus of the ADBC program is on digitization and data production (and not research), the ADBC program has been working with other NSF programs to call attention to data needs. For example, the Advances in Biological Informatics (ABI) program in DBI, which supports development of new tools for biology, already has received several proposals for developing tools for the research community that utilize collections data. Additionally, iDigBio has conducted outreach activities at all major research meetings, has been in discussion with scientists at NEON about integration, and has been present at principal investigators (PI) meetings for the Dimensions of Biodiversity program and the Assembling, Visualizing and Analyzing the Tree of Life program. The MSB program requires investigators to provide detailed data and project management plans as part of the proposal submission. Several MSB projects are developing computational tools, such as data assimilation to incorporate streaming data into models, algorithms, and methods to conduct multi-model analyses. The MSB program will continue to encourage and support further development of such tools.

I. Questions about the quality and effectiveness of the program's use of merit review process. – No Recommendations

II. Questions concerning the selection of reviewers. – No Recommendations

III. Questions concerning the management of the program under review.

Recommendation 4 (III.1, page 8): The dwell time for some EF proposals is excessive and should be reduced; for at least one cycle proposals took 6-9 months for a decision to be transmitted to PIs. The

program staff members have already considered this and think it may be possible to speed up the process with a simple change of due date and related panel dates to increase availability of panelists. We recommend that this be done.

Response: The long dwell time for the MSB program is a function of having a deadline (April) in one fiscal year but funding occurring in the subsequent fiscal year. The panel is held in September with award decisions occurring in the next fiscal year. Delays in award processing also have been exacerbated in the past few years due to uncertainties in the federal budget that have been unresolved until late in the fiscal year. In some years, those uncertainties were not resolved until the summer, almost nine months after program recommendations were finalized. The MSB program will make every attempt to shorten the dwell time.

Recommendation 5 (III.2, page 8): Many of the proposals, especially for MSB, introduced substantial innovation with respect to the management and governance of large multidisciplinary efforts; this itself seems to us an important Criterion 2 activity that might be pointed out to reviewers.

Response: BIO concurs with this recommendation.

IV. Questions about Portfolio.

Recommendation 6 (IV.2.d, page 11): The awards made by the ADBC program include more than 200 institutions, largely at PhD granting institutions, but including commercial and NGO entities. As the digitized collections created under program funding come on line, there will be ample opportunities for other disciplines to utilize digitized collections. The committee would like to see additional plans for collaborating with federal collections.

Response: Please see the response to Recommendation 2. Plans are underway and will continue as the federal collections begin to digitize their holdings. Collaboration is currently ongoing. When data is available for integration from the federal collections, it will occur automatically through BISON.

Recommendation 7 (IV.3.a, page 11): The proposals received by MSB have been generally responsive to the solicitation. With the completion of several cycles, it now seems appropriate to revisit the solicitation to ensure it is aligned with the current aims of the program as reflected in funded proposals.

Response: BIO concurs with this recommendation. The MSB program will be revising the solicitation in the next 6 months and incorporating recommendations by the COV.

Recommendation 8 (IV.3.b, page 11): As the [MSB] program matures the foundation should consider reaching out to NASA, USGS, DOE, and other such agencies for partnership opportunities.

Response: BIO agrees with the COV that this is an important endeavor. The MSB program has invited program officials from the National Aeronautics and Space Administration (NASA), USGS, and the Department of Energy (DOE) to participate in panels as observers. The program will continue to do so in future panels and reach out to these agencies for partnership opportunities.

Recommendation 9 (IV.3.c, page 12): The [MSB] trainingship awards are also too early and too few to judge their ultimate success. These awards will likely be very important to build workforce capacity towards the ability to engage in macro-scale questions. The program is urged to consider a targeted call for proposals based on input from the community about emerging needs. The program is also urged to consider EHR as a potential funding partner for trainingships, as well as other agencies.

Response: The MSB program will continue to work with the community to develop greater awareness of training needs. It will strengthen the training component to highlight the importance of training to the MSB program in the upcoming revisions to the solicitation. Furthermore, the program officers will strive to promote the program's interest in innovative training by organizing additional outreach activities in the relevant communities at national meetings (e.g., town halls and open houses) and by providing materials to program officers in other offices within NSF such as EHR. The program will seek additional insights and perspectives from EHR as appropriate to improve training opportunities.

OTHER TOPICS

Recommendation 10 (1, page 13): Given the evolving nature of both the MSB and ADBC programs, it is time to revisit the program descriptions to ensure they are aligned with proposals that are successful and to integrate solicitation modifications into the body of the solicitations themselves. In both cases, including pointers to additional sources of program information will be helpful. In the case of ADBC proposals, it would be helpful to specifically include additional information about the hub site (iDigBio) so as to encourage growing community use of its resources.

Response: BIO concurs with this recommendation with the caveat that we do not want the current portfolio of either program to limit the submission of future innovative or "risky" new proposals. The program will review the portfolio composition, annual reports, project publications, and recommendations for the annual PI meetings to assess the alignment with program descriptions in the solicitation. This analysis will inform the upcoming solicitation revisions. In addition, we will add resources as relevant to the solicitations. For ADBC, it is important to note that for most new solicitations for NSF programs, iDigBio is listed as a resource (see solicitations for Dimensions of Biodiversity, Genealogy of Life, and GEO programs under EarthCube). Additionally, iDigBio is constantly updating its resources in response to community needs. The solicitation points to the iDigBio website as a source for information regarding the resources offered. The ADBC solicitation is updated every year and we will continue to do so, as recommended by the COV. For MSB, the revised solicitation will include links to the Frontiers in Ecology and Environment Special Issue.

Recommendation 11 (1, page 13): Additional interaction/integration between the ADBC and MSB programs might be useful to identify and solve common data management challenges.

Response: BIO concurs with this recommendation. The programs have begun to explore connecting their PIs and will consider referencing one another when revising solicitations.

Recommendation 12 (1, page 14): For MSB proposals, increased emphasis on accountability for management of data, models, etc. in the annual report system for Category II proposals is important for both ensuring that commitments are met for providing BIO information about what PIs are doing with respect to information management, what is effective, and where problems remain.

Response: BIO concurs with this recommendation. The MSB program plans to add an additional requirement in annual reports submitted by PIs to provide status updates on data management and project goals. The PIs will be notified of this new requirement and a discussion about this proposed requirement is planned to occur at the upcoming PI meeting. This change will also be included in the revised solicitation.

Recommendation 13 (3, page 13): A joint effort to develop interoperability, community governance and other capabilities that are general to environmental data management broadly would be a highly valued effort.

Response: Please see response to Recommendation 1.

Recommendation 14 (3, page 13-14): It also seems time for NSF to begin an interagency collaboration relating to collections. The fact that iDigBio is now operating, and that numerous collections are being funded to digitize and deposit their data in iDigBio provides strong support for such an interagency approach, including digitization of federal collections and the expansion of the central database. The COV also felt that it is also time for NSF to initiate a plan to develop the tools that will be required for utilizing the massive specimen database to its full potential.

Response: Please see response to Recommendation 1 and 2.

Recommendation 15 (4, page 14)The committee feels it is essential to make sure there is a long-term home for integrative, highly collaborative science, whether it is MSB or some evolution from MSB, is identified.

Response: BIO concurs with this recommendation and will continue to discuss this issue within BIO.