

2018 Report of the iDigBio External Advisory Board

November 21, 2018

iDigBio External Advisory Board (EAB) membership:

- Neil Cobb, Northern Arizona University (Chair)
- Linda S. Ford, Harvard University
- Donald Hobern, Global Biodiversity Information Facility
- Eva Huala, Phoenix Bioinformatics
- Mary Klein, NatureServe
- Jason Knouft, Saint Louis University
- Thomas Orrell, National Museum of Natural History
- Vince Smith, Natural History Museum, London
- Barbara Thiers, New York Botanical Garden

This report is based on the following:

- NSF site review and the response by iDigBio
- iDigBio “Committee of Five (CoF)” report
- In-person meeting with all EAB members, iDigBio key personnel, along with Muriel Poston and Roland Roberts (NSF observers) on November 1, 2018 at Summit VII in Gainesville, Florida (minutes included as **Attachment 1**)
- Presentation by Gil Nelson at EAB in-person meeting (November 1, 2018)

The EAB is charged with providing recommendations to iDigBio regarding sustainability, with consideration given to operating after the end of the NSF-ADBC program, and more specifically increasing data use by research as well as education-outreach activities. In 2018, the EAB worked as a single committee to assess sustainability and core functions of iDigBio, including data use.

The EAB reviewed two primary documents, the NSF site review and the Committee of Five (CoF) strategic plan that was completed in August 2018, following the 2018 NSF site review and iDigBio response submitted to EAB in July 2018. The 2018 EAB Report focuses on recommendations regarding the CoF report and presentation as well as assessments of issues identified in the 2017 EAB Report.

General EAB Recommendations

Below are the key points based on the 2018 meeting at the ADBC Summit.

- 1) The EAB appreciates the leadership exhibited by Larry Page in guiding iDigBio through its first eight years, and fully supports Gil Nelson, who will replace Larry Page as the director of iDigBio.
- 2) During the last year, iDigBio has developed a series of ambitious planning documents for post-ADBC operations and funding. The cumulative material is comprehensive and provides an excellent basis for long-term strategic planning. Most of the EAB recommendations emphasize further refinement and prioritization of these plans outlined in the Committee of Five (CoF) document and presentation.
- 3) The EAB suggests that iDigBio frame the action items listed in the CoF report in the context of the (seven) "critical functions" or any refined version of these.
- 4) We encourage iDigBio to refine the "critical functions" as key services offered to the community or as user stories rather than as broad areas of function. Furthermore, iDigBio should present a business plan and sustainability strategy for each such service, recognizing that the strategy may be for collections and/or other parties to assume responsibility for some of these services in the longer term. iDigBio should identify an achievable number of core functions to be sustained by iDigBio itself.
- 5) International partnering leading to a globally connected pipeline will be the most significant and challenging aspect of iDigBio planning.
- 6) We are confident that iDigBio will accomplish all the goals set for the original 10-year period and is well positioned to complete its ADBC directive.

In summary, the EAB supports the direction and planning activities initiated by iDigBio with the ADBC Strategic Implementation Plan. We strongly encourage the iDigBio to continue these efforts with more specific plans for how the goals of an "ADBC" effort will be sustained after 2021, with careful consideration for prioritizing functions and collaborations both with the ADBC community and beyond to create a global biodiversity pipeline.

Specific Recommendations on CoF Planning Document & 7 Core Functions

The Committee of Five Document and planning Phases

The CoF document is titled “ADBC Strategic Implementation Plan”, but the document appears to focus on iDigBio activities. If this plan encompasses all of ADBC, rather than just iDigBio, it would be helpful to see more detail on how iDigBio would engage the collections’ institutions in developing and “owning” the long-term, sustainability strategy. There are currently items listed that relate to supporting the data systems (e.g., Symbiota, SCAN, Specify, Arctos) and existing initiatives (e.g., data carpentry, Darwin Core Hour), but much less specificity around institutional engagement.

The structure of the CoF document is focused on processes and mechanisms rather than goals and deliverables. We suggest that the ADBC Strategic Implementation Plan should instead be structured around a clear view of a set of services that iDigBio should assure for the long-term benefit and support of the collections community.

The Seven Critical (i.e., Core) Functions identified in the presentation delivered by Gil Nelson at the EAB meeting (listed below) offer an effective framework for presenting these services. However, the EAB believes that the critical functions as currently listed are too generic, could probably be restructured as a smaller set, and should be more tightly defined as a set of key services that iDigBio is offering to the community. It may be most effective to present these services as a set of user stories – e.g., "A collection manager uses service A from iDigBio to achieve X, Y and Z". Such a view of the iDigBio offerings would give clarity and confidence for all stakeholders about the current scope and vision.

Critical Functions and Planning Phases

iDigBio Core Functions

1. Workforce training/community building
2. Education, outreach, and diversity (EOD)
3. Data mobilization
4. Data support/troubleshooting
5. Special access to specimen data
6. Encouraging/tracking research use of data
7. Continuous program evaluation/documentation

iDigBio Planning Phases

Phase 1: 2019 to 2021 (August)

Phase 2: 2021 to 2024 (post ADBC funding)

Phase 3: Post 2024

Defining this set of key services, which should focus on the capabilities, both technical and social, that are being supported by iDigBio during the current program, will allow the ADBC Strategic Implementation Plan to offer a future sustainability vision for each of these services. In some cases, the future plan may be to secure adequate resources within iDigBio to maintain the service at the level needed by the community. In other cases (e.g., some aspects of workforce development), the goal may be to ensure that the collections community itself has the capability to maintain the service following the end of the current funding period. In yet other cases, the sustainability plan may reduce long-term resource needs through planned closer collaboration with GBIF, ALA or other external parties to ensure cooperative provision of a shared service.

An ADBC Strategic Implementation Plan of this nature could support fruitful discussion with collections, with other research infrastructures, with NSF and with other funding sources to secure an optimal approach to sustainable digital management of US collections. This approach can also clarify iDigBio's unique niche as the catalyst for coordinating efforts and resources to ensure that researchers have the capability they require.

The Plan can also present, for each service, a timeline for how each service is to be maintained and improved over Phase 1 and Phase 2. By the conclusion of Phase 2, iDigBio and the ADBC community should be in a maximally sustainable form for Phase 3. Regarding Phase 3, iDigBio should provide a general perspective for the period around 2024-2028, regarding what future investments might possibly build on the foundations of the ADBC program and how these might either depend on the sustainability of the critical iDigBio services or else host components of the next phase of a persistent iDigBio capability.

The 34 action items listed in the CoF document should be linked to the critical functions. They could be further developed in more detailed plans defined by the seven critical functions. For each of the services identified as necessary on behalf of the collections community and researchers through 2024 and onwards, the focus should be answering several important questions, including:

1. What is the plan, and what resources are required, for maintaining and improving this service during Phase 1?
2. What can be done to maximize the sustainability of the service by the end of Phase 2?
3. What are the options for provision of this service in Phase 3? This raises a suite of questions the EAB feels are critical for long-term sustainability of the ADBC effort.
 - A. Does this service need to remain part of a continuing iDigBio/post-iDigBio centralized infrastructure?
 - B. Could it be maintained sustainably as part of the operations of collections that have benefited from ADBC?

C. Is it possible that partners such as GBIF can provide redundancy to guarantee persistence, or some or all services be transferred to an interested institution on an interim or permanent basis?

The Internationalization section is about a strategic approach to collaborate on developing a global pipeline. This would involve collaborating with GBIF, ALA and other programs and allows iDigBio to leverage existing efforts more effectively. This has been the area where iDigBio has made the most progress in terms of developing a strategic plan for specific collaborations with GBIF, ALA and other programs. Developing closer ties with the international community is likely to reduce redundancy allowing more focus on other areas that will lead to a stronger cyberinfrastructure better able to serve researchers the data they need. The contents of the Internationalization section then become an explanation of plans during Phases 1 and 2 to evolve some of the existing iDigBio services to a more persistent and sustainable shared model. The reason and justification for developing partnerships is to enhance iDigBio infrastructure while also improving its sustainability. Partnering internationally is the proposed mechanism to do this and the “Internationalization” section is good but should be re-named to better fit iDigBio and ADBC goals.

The remaining two sections of the CoF document appear to be more detailed action items that encompass the core functions. The second section in the CoF (Education, Outreach, and Diversity (EOD)) is an expansion of the first and second “Core Functions”, and section III (Preservation) is about sustaining Core (i.e. Critical) Functions.

Seven Critical Functions

The EAB members suggested several ways of integrating the seven core functions. Specific possibilities included combining #3 (Data Mobilization) and #4 (Data support/troubleshooting), and both of these fit well with #6 (Encouraging/tracking research use of data). The core functions #1 and #2 could be combined into a 'Training, education, outreach focus'.

Critical function #1 (Workforce training/community building), #3 (Data Mobilization) and #4 (Data support/troubleshooting) are the most fundamental of functions; furthermore items 3 & 4 can be viewed as different aspects of the same function. Additionally, without number 6, the data mobilization effort doesn't have any impact and/or its impact can't be demonstrated, thus it might be appropriate to integrate core function # 6 (Encouraging/tracking research use of data). Funding for iDigBio, partnerships (with TCNs, scientific societies, GBIF/ALA, educational initiatives, federal agencies and others), EOD, and new funding proposals all become activities that can be prioritized and pursued based on their likely contributions to sustaining, growing and evolving the essential functions. We do not feel that core function # 7 is necessary to delineate as a separate function, but could be acknowledged in the other core functions. The EAB is not

requesting the elimination of any core function; it would likely be premature to do so without additional rounds of planning.

There is some concern that even the “Seven Critical Functions” are too general. These are areas of activity rather than defined services or capabilities that iDigBio offers to the community. It could be much more powerful to define the key services that iDigBio offers to its stakeholder community, perhaps defined as, or at least clarified by, user stories (as noted earlier). The list of core functions should provide a sharper, crisper set of definitions of what iDigBio delivers for each of these separately. For example, what are the critical functions of iDigBio concerning Data mobilization (#3)? Which of the following are included in iDigBio's critical function: documentation, hands-on support, helpdesk, training courses, publishing tools, data hosting services? Is #4 (Data support/troubleshooting) actually a separate function from #3 (Data mobilization)? For some of these, the need to clarify is even greater - especially EOD.

Once there is a clear view of the actual services that iDigBio could maintain going forward, it should be possible to rank these into Critical, Desirable and Optional. This type of ranking would offer more flexibility as funding potentially varies, and it demonstrates a priority analysis of the functions that could be provided to the collections community. At this stage, it may be wise to identify the most essential core of iDigBio rather than committing to do everything indefinitely.

Attachment 1: Minutes from the 2018 iDigBio Meeting with External Advisory Board

Date/Time: Wednesday, November 1, 2017, 1:00-5:00 PM Eastern
Location: McGuire Center Director's Conference Room, 2nd Floor, Florida Museum of Natural History

Attendees

- **EAB:** Mary Klein, Barbara Theirs, Linda Ford, Neil Cobb, Jason Knouft, Donald Hobern, Vince Smith, Eva Huala
- **iDigBio:** David Jennings, Gil Nelson, Renato Figueiredo, Bruce MacFadden, Pam Soltis, Jose Fortes, Shari Ellis, Larry Page, Greg Riccardi
- **NSF observers:** Roland Roberts, Muriel Poston

EAB meeting began with a presentation from Gil Nelson

https://www.idigbio.org/sites/default/files/internal-docs/eab/2018-10/Summit_2018_revised.pdf

Sustainability of ADBC: Facing the potential for sunset in 2021, the iDigBio Executive team established the Committee of Five to develop an implementation plan leading to the continuation of what ADBC had started (Shari Ellis, Molly Phillips, Renato Figueiredo, Gil Nelson, and Joanna McCaffrey).

The CoF Interviewed: Larry Page, Pam Soltis, Kevin Love, Greg Riccardi, Bruce MacFadden, Jose Fortes, Doug Jones, and Donald Hobern,

They identified seven critical functions of iDigBio:

1. Workforce training community building
2. Education, outreach and diversity
3. Data mobilization
4. Data support troubleshooting
5. Special access to specimen data
6. Encouraging/tracking research use of data
7. Continuous program evaluation/documentation

With a 3-Phase plan

- Phase 1: the next 3 years
- Phase 2: years 11-13 (post June 30, 2021)
- Phase 3: Beyond year 13

Phases 1 and 2 are beginning to fade into each other. We still need to figure out what will happen after TCN funding ends. The CoF drafted an implementation plan as a way to plan out each phase of the sustainability of iDigBio.

Internationalization

A meeting at GBIC2 occurred among the GBIF, ALA, and iDigBio programs to talk about how to work together to remove barriers among the national silos. How can we combine our efforts into one major data

store? Our sustainability is dependent on working internationally. Now we are putting together a sandbox/pilot to further develop how the 3 projects can work together.

Each project has unique strengths and some overlap for synergy

GBIF

- Global leader in data aggregation and mobilization
- Worldwide organization and network

ALA

- Provides analytical and visualization tools for biodiversity research
- Global collaborator

iDigBio

- Leader in data mobilization and visualization
- Expertise in digitization, community development and workforce training

iDigBio would remain focused on mobilizing data in the US (as a US node) but start operating as a global collaborator through the GBIF framework. Next steps are integration of projects.

Some of the practices we will need to integrate:

- Identifier synchronization
- Data quality integration
- API integration
- Tool development
- Data integration
- Data ingestion alignment
- Collections catalog maintenance
- Citation/attribution protocols
- Workforce training
- Digitization practices
- Gap analysis

Ultimate goal is a single source store for biodiversity data...

The plan is to continue as a GBIF participant under the iDigBio moniker while exploring avenues of continuing support for mid-level data aggregators. Also plan to explore avenues of continuing support for ADBC-related collaborations.

In about three years, we will transition the physical presence of iDigBio at UF into new facilities of the Florida Museum of Natural History while maintaining the existing strong collaboration with ACIS and the iDigBio office at FSU.

Diversity

- Provide leadership for expressing diversity as a core value
- Engage with TCNs to reach underserved communities
- Train those communities with the skills and perspectives required to work with digitized collections
- Disseminate best practices
- Model a commitment to inclusiveness
- Collaborate with Minority Serving Institutions (MSIs) currently involved in ADBC/collections

Human Resources Security

Implement retention and recruitment plans to address critical continuing skills needed for leadership and technical staff to ensure maintenance of institutional wisdom...

Discussion Portion of the Meeting

GBIF and iDigBio collaboration is similar to collaborations that formed among genetic/sequencing facilities, so one of the first suggestions from the EAB was to arrange meetings with these organizations to get lessons learned from the success of other communities.

The EAB wanted to know how we are going to divide up tasks and leadership roles among the different programs. The response was that many issues need to develop collaboratively, such as data standardization and identifiers. Ultimately we need to work together to solve the issues that are left. As an example, it does not matter who assigns a DOI, it does matter if all of the programs are operating with the same standards and best practices when assigning DOIs. A large portion of the Summit will be dealing with how we should be moving forward collaboratively.

Another concern raised by the EAB was the amount of goals and bullets point included in the implementation plan, with the worry that iDigBio is trying to sustain too much. The response was that this is the first draft of the document and the iDigBio team imagines that there will be streamlined and consolidated. They also plan to create several plans – including one that identifies the minimal levels of sustainability of efforts.

It is also really important to identify community members to collaborate with. Many TCN members are just as invested in sustaining the national digitization effort as the iDigBio team. iDigBio should also work collaboratively on the implementation plan goals with GBIF and other partners. GBIF for instance, is also looking at risk assessment and create worst case scenario cases for their nodes. GBIF is thinking about how to do a stakeholder analysis so we could work together on that goal too. This could help unravel the connections and roles between all of the different entities as well.

The iDigBio team hopes to have a finalized draft of their implementation plan in the next six weeks.

One of the EAB members commented that they thought that iDigBio's site review response was too general, but the implementation plan was a much more fleshed out response.

Until the final plan is submitted and Roland Roberts has had a time to review it, he can't respond to the likelihood or degree of future funding that could come from NSF for any portion of the sustainability implementation plan.

There was discussion on the distinction of the roles between BISON vs iDigBio. BISON is still the US Node for GBIF. People have discussed the idea of iDigBio taking over the North American Node for GBIF. However, there is no such thing as a regional (like North American Node). GBIF has country nodes and other projects that can be considered nodes (like BHL). The term "node" was loosely defined as a data source in the beginning of GBIF, but has morphed into meaning more of a networking unit. GBIF could split and divide our network partners in any way that makes sense. The definition is not fixed, however, we need to better define the roles moving forward. One thing to consider is that it may be easier to rally the community around a national database if it was a single entity (instead of having BISON and iDigBio). GBIF, NSF, iDigBio, BISON, and USGS should sit down and figure out the

semantics of reach project's role to satisfy everyone's needs. Becoming an international partner is crucial to ADBC's continued relevance and our best shot for additional funding.

Thinking about the sustainability of phase 2: ADBC was promised funding for 10 years. NSF intends to continue funding TCNs and PENs into 10th year. Conversations have begun within the NSF about how to sustain iDigBio through at least the duration of the last TCN awards.

Donald Hobern asked for clarification on the tone of the site report, because he had a hard time reading the overall impression from the site visit. The response from the NSF representatives present was that the site visit panel was asked to pay specific attention to certain topics -- the topics that were focused on in the report. Not all the recommendations are in scope, either. Some of the recommendations will be expanding iDigBio's activities instead of streamlining with the focus of sustainability. There is flexibility at NSF for iDigBio to push back on some of the recommendations from the site visit and argue that it is out of scope or not sustainable.

One note about sustainability of the effort through the current funding source is that the NSF DBI budget has stayed the same since 2008. We need to consider fee structures, and seeking support through other agencies and other sources of support outside of the NSF. International collaboration strengthens the justification for funding with national funding organizations.

The EAB would like to get more detail on the Minimal Viable Product for the collaboration among ALA, GBIF, and iDigBio and see also see some tightening up of the definitions of the different roles iDigBio and BISON will play moving forward as part of GBIF. Finally they, would like to see some reduction and consolidation of the goals in the implementation plan.

The meeting concluded with a discussion of whether iDigBio should embrace data beyond collections – like data from iNaturalist. People pointed out that iNaturalist is middle ground between embracing all citizen science data because iNaturalist records at least have photo “vouchers”. Many other observational data sources have no “voucher” or evidence. Others in the meeting argued that adding new data sources is once again expanding and not contracting and streamlining iDigBio efforts. We could at least poll the community and see how they feel about incorporating observational data. iDigBio has become a global center of expertise for museum digitization. Broadening the scope, may lose what is special about the program, and create overlap with other projects. We are also not even near the goal of digitizing all the specimens in the US, maybe we should not be getting into mobilizing other data sources. It was not the original goal to digitize everything. Original goal was to digitize subsets of museum data for research topics – hence the creation of TCNs and PENs.

iDigBio could be partnering with other entities to work more on citizen science and other public engagement (and other EO&D activities) instead of taking on these efforts on their own. Museums already have great community engagement so they would be some obvious partners – FLMNH along with institutions across the ADBC network.

iDigBio could work on developing materials for different audiences with different collaborators. However, many TCNs and PENs already making these types of materials, so iDigBio should just continue to help promote the materials already being created through ADBC.

iDigBio is not expected to do everything in a vacuum, but is expected to leverage the community and coordinate these efforts.

How do we need to keep the TCNs engaged? How should we encourage TCN PIs to be more involved and have more ownership in ADBC's future? iDigBio is infrastructure which is why NSF is interested in sustaining the bub and note the TCNs. Theirs clarified she does not mean sustaining the TCNs as they are, but more engaging the institutions that house some of the more engaged TCN PIs. Institutions that have collections have some of the most investment in sustaining the national effort.