2014 Report of the iDigBio External Advisory Board

2014 iDigBio External Advisory Board (EAB) membership:

- Linda Ford, Harvard University
- Donald Hobern, Global Biodiversity Information Facility
- Paul Kimberly, National Museum of Natural History
- Mary Klein, NatureServe
- Vince Smith, Natural History Museum, London
- Barbara Thiers, New York Botanical Garden

During the iDigBio Summit IV in Gainesville, Florida, 27-28 October 2014, the EAB reviewed progress and directions within iDigBio. This review included the following:

- Attendance by five members of the EAB (all excluding Vince Smith) in plenary and working group sessions of the Summit
- Discussion of all EAB members (Vince Smith by telephone) with iDigBio PIs and staff and NSF Observers, 27 October 2014 (minutes included as *Attachment 1*)
- Closed discussion by all EAB members, 27 October 2014
- Closed discussion by five EAB members (all excluding Vince Smith), 28 October 2014

Progress with iDigBio implementation

- EAB members commend iDigBio for progress during the first phase of the project, and in particular during the past year
- iDigBio workshops, working groups and digitization resources are providing significant leadership and coordination for mobilization and use of US natural history collections
- iDigBio has made significant progress in integrating species occurrence data and multimedia resources through the iDigBio portal
- Updates from TCNs during the Summit showed that the TCN model is serving to foster an exciting range of active and expanding communities with interesting and valuable research focus
- TCNs are successfully mobilizing and organizing significant rich specimen-based datasets with many additional properties beyond those accessible through the iDigBio portal

EAB Recommendations

The EAB makes the following recommendations for future development and sustainability of iDigBio. We indicate whether iDigBio should address each matter as a priority in the short term (within 3 years) or as part of a long-term strategy for growth and sustainability (within 10 years).

Vision and strategy

Short term recommendations (within 3 years)

- iDigBio must clarify its unique niche within biodiversity research, biodiversity informatics and research infrastructure. The EAB strongly urges iDigBio to undertake a stakeholder analysis, in particular to address the following questions:

    - Which communities have a need for specimen data mobilized through iDigBio?
    - Which of these communities should be served directly through iDigBio tools and which can be served via other intermediaries (which should themselves be identified as key consumers of iDigBio products)?
    - In what form do these communities require data to be delivered (raw data from collections; normalized curated views; summary metrics and indicators; infographics; geospatial products; etc.)?
    - Which other stakeholders might serve as alternative or complementary channels for use of data from iDigBio?
    - How should the mission of iDigBio relate to these other stakeholders and how should unnecessary duplication and competition be avoided?
    - What are the responsibilities, tools and services which are part of iDigBio's core mission (and which only iDigBio is positioned to deliver)?
    - What secondary areas should iDigBio address to maximize the effectiveness of its delivery in core areas?
    - How does iDigBio build partnerships to avoid gaining responsibility to maintain an unsustainable number of products and services?

- iDigBio should in particular urgently clarify its role and responsibilities in relation to BISON. Is BISON a customer, key partner, or channel? The current division of responsibility appears to be based on a division between federal and non-federal data sources and on iDigBio mobilizing globally relevant content rather than purely addressing national perspectives. Present plans seem to allow for significant duplication in effort, particularly around the presentation of data for spatial planning and conservation, and a need for bidirectional data exchange to ensure data completeness. A key use of specimen data is to support spatial planning and conservation, and iDigBio certainly needs to demonstrate the value of its data delivery in supporting these requirements. However, iDigBio could rely on BISON or other channels to deliver appropriate tools (including integration of relevant non-collection data, which needs to happen but which seems to be outside iDigBio's mission) and could itself focus on high-value data management for rich integrated collection data (see more below). The EAB encourages iDigBio to distinguish between the value of its work for a particular application and the degree to which iDigBio itself needs to own delivery for the application.

- iDigBio must also urgently clarify how it will work with GBIF. The EAB believes that synergies and expected linkages between iDigBio and GBIF are clear, but note that several Summit attendees raised questions on the relationship and particularly on whether data resources

needed to be separately registered with iDigBio and GBIF. It is important to clarify data pathways and collaborative opportunities with GBIF. Similar clarification may also be needed between iDigBio and other data integration networks such as VertNet and the Encyclopedia of Life.

- In general, the EAB emphasizes that there should not be a radical departure from the current focus areas for iDigBio. Significant effort is still required in the following areas:
    - Changing the culture of institutions so that digitization becomes a normal routine element within everyday collections activities.
    - Engaging with ALL US natural history collections including the Smithsonian as partners in mobilizing specimen data.
    - Developing and promoting best practices for collection digitization.
    - Providing training and support for collections staff in digitization and data management procedures.
    - Adopting and promoting data standards to enable collection managers and taxonomists to standardize rich and consistent content for each taxonomic group (extensions to Darwin Core properties, standardization of vocabularies, mobilization of multimedia, sequences, morphometric data, taxon interaction data, etc.).
    - Providing rich integration of all data from existing and future TCNs and from taxonomic or regional networks.
    - Providing tools and best practices for rapid establishment of new TCNs and other networks.
    - Developing robust long-term repositories and research infrastructure tools for preservation, curation, enhancement and use of iDigBio data.
    - Contributing to the development of a research culture based on citation and recognition for contribution of open data.

- Consequently, the EAB urges caution around expansion of iDigBio effort into new areas. Linkages with education, conservation, citizen science and other user communities may increase impact and enhance understanding of needs. This may in turn lead to greater sustainability. However, as far as possible, iDigBio should ensure that partners with responsibilities in these areas assume responsibility for development and delivery of any resulting products.

- In regard to digitization through TCNs, the EAB is concerned that priority may have been given to lowest-cost solutions rather than to ensuring that digitization efforts are not only efficient and cost-effective but also capture sufficient data elements and multimedia to support user needs. It is important for iDigBio to understand and respond to the actual data needs of expected users and this should be taken into account when planning any ADBC digitization. Consideration should also be given to likely future costs if the same specimens require further digitization work at a later date. The EAB recognizes that iDigBio does not have responsibility for TCN funding decisions, but believes that iDigBio can influence this by continuing to define best practice for specimen digitization for different taxa, by contributing to standards definition for

taxon-specific data elements and by integrating all data elements from TCNs through a single portal.

- It should be possible to explore relationships between specimens of host plants, herbivorous insects and their parasitoids within the iDigBio portal. The portal should expose a rich graph of interconnected data elements - specimens, sequences, trees, etc. A focus on rich data integration, rather than simply on data management to address TCN research needs, will highlight the opportunity for iDigBio to develop as a virtual natural history collection supporting modern taxonomy.

- iDigBio should work with NSF to require funded collections to share relevant data through iDigBio, and to work with TCN PIs on digitization compliance issues.

- Limited guidance was supplied to EAB members on their roles and responsibilities. For next year's Summit, the EAB should be provided with a robust agenda and necessary documents ahead of time.

Long term recommendations (within 10 years)

- There needs to be a gradual shift towards establishing an achievable sustainability strategy - especially with an eye towards long-term data management, training and facilitation of the community of practice. This will only be achievable if iDigBio has a well-defined and unique niche. The EAB emphasizes the importance of developing sustainable funding and governance models at the earliest possible opportunity.

- ADBC lacks a component with an explicit responsibility to coordinate taxonomic expertise from throughout the collections network. iDigBio has responsibility only for content integration. The sustainability and quality of these data will depend on continued and deepening engagement with collection managers and taxonomists to use iDigBio services and to correct, update and enrich data. This implies the need for iDigBio to consider expanding its vision to becoming an expert community network as well as a content network, probably in collaboration with one or more professional societies. A network combining content and expertise would be well positioned to provide high quality syntheses of knowledge and would therefore be likely to address more of the needs of users of biodiversity information.

Attachment 1 – Minutes from EAB meeting, 27 October 2014

Meeting Date/Time: October 27, 2014 / 4:15-5:30 PM EDT with an EAB-only session from 5:30-6:15 PM EDT

Attendees:

- **External Advisory Board**: Vince Smith, Donald Hobern, Paul Kimberly, Barbara Thiers, Linda Ford, Mary Klein
- **iDigBio PIs**: Larry Page, Bruce MacFadden, José Fortes, Greg Riccardi, Pam Soltis
- **iDigBio Staff**: David Jennings, Shari Ellis, Austin Mast
- **NSF Observers**: Roland Roberts, Judy Skog

Adobe Connect Recording: http://idigbio.adobeconnect.com/p95crp8ve76/

Action Items

- (David Jennings) Provide the EAB with copies of iDigBio's Cooperative Agreement and FY3 Annual Report including copies of the annual evaluation and last year's EAB report. *[Completed on 10/28/2014]*

Meeting Minutes

Role of the External Advisory Board

"External Advisory Board (EAB) whose membership will be subject to the approval of NSF's cognizant program official and will meet at least once a year; provide written and verbal advice to the national resource on its activities, including progress and integration of digitization projects, research, education and outreach activities among all funded institutions; advise the national resource leadership on strategic directions and management policies."

iDigBio Progress Report

An overview of iDigBio's progress by staff was not provided during the EAB meeting. The meeting occurred during iDigBio's Summit IV where everyone heard about iDigBio's activities through presentations, posters, and demonstrations. iDigBio feels that having the EAB meeting at the Summit is an effective way to communicate its activities and collaborations.

iDigBio intends to continue conducting EAB meetings annually at the Summit unless the EAB objects. No objections were raised.

EAB Report

iDigBio reminded EAB that the EAB report is due to NSF within 21 days of the meeting. The EAB agreed, which precipitated an EAB-only session at the end of the meeting.

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Specific advice from the EAB to iDigBio

iDigBio is in year four of the original 5-year grant and will be submitting a proposal for renewal next year.

iDigBio is particularly interested in what activities the EAB thinks iDigBio should pursue in the renewal years 6-10.

Questions & Responses

Is iDigBio planning to stay the course, or does iDigBio have more than one phase in mind for the project?

- Phase one (years 1-5) has been focused on data mobilization.
- Phase two (years 6-10) will transition to a focus on data uses in education and research.
    - iDigBio envisions that the next few years will be push-and-pull with the collections community.
        - iDigBio now has enough data in the portal to identify research applications, but iDigBio expects this to result in some push back from the community to improve and expand the data.
        - iDigBio will attempt to pull more (and better quality) data from the collections community.
    - iDigBio will need to provide data and training for research and actively promote research.
    - Ultimately, iDigBio's goal should be to address higher-level questions with collection data.
    - iDigBio wants to develop research capacity outside its current network and collaborate with other groups.
- Other foci mentioned:
    - Sustainability
    - Continue connecting the collections community through workshops and other events ("social sustainability").
    - Getting data for the portal from other sources (including international).
    - Education and Outreach
        - K-12 and undergraduate education
        - Increasing public participation
            - Examples given included fossil clubs and hackathons (CitScribe and CitStitch).

Would focusing on research mean less focus on digitization?

- Years 6-10 will still contain workshops and support for digitization training because there is still need in the collections community for digitization support.
- As TCN projects mature, iDigBio will move towards data uses and providing training and support for research and education.

What are iDigBio's plans for participation in K-12 and undergraduate education?

- iDigBio would like for years 6-10 to have a K-12 outreach component.
    - iDigBio wants digitization to become a household term.
    - New science standards make iDigBio's topics easier to integrate into K-12 curricula.

- iDigBio is in the process of hiring an Education & Outreach Coordinator to help develop education and outreach programs.

How is iDigBio planning to connect with conservation use?

- Conservation is one of the rationales for the ADBC program.
- IDigBio involvement in policy-making needs to be approached with care, but conservation uses for specimen data is a direction that iDigBio should move towards.

Is promoting and conducting conservation planning and research iDigBio's niche? Would there be a duplication of effort from other entities?

- iDigBio wants to avoid duplication of effort, but there is a high level of expertise in iDigBio that should be exploited.
- iDigBio could be an organization that informs conservation policy as one of its roles.
- Collections are the most reliable and verifiable sources of information for species distributions.

How does iDigBio avoid "ownership" of services (i.e., iDigBio is responsible for running services for the long term)?

iDigBio could stick with providing infrastructure and tools for TCN research (and other endeavors) as its focus.

Does iDigBio really know who its customers are?

- There was agreement that it would be a useful exercise to identify iDigBio's customers.
- iDigBio needs to reach out beyond traditional uses of specimen data.
- Field biologists are an untapped resource because they are interested in expertly identified photographs for use in identification in the field.

When does iDigBio become recognized as "THE place to go"? When will iDigBio become the authority for specimen data?

Eventually iDigBio will reach a tipping point where everyone who has a collection knows about iDigBio, and all researchers will cite iDigBio when using collections data. **This should be a goal of iDigBio.**

Is there funding to promote use of specimen data (directed to NSF observers)?

- NSF is providing multiple methods of support for collections-based research:
  - New Research Coordination Network will examine ways to sustain the national digitization network.
  - NSF Challenge Award for imaging insect trays.
  - New postdoctoral awards for research utilizing collections (innovative research and mentoring for future careers).
- NEON has a biodiversity component.
- Any project on biodiversity that receives funding from NSF should be talking with iDigBio.

Do you expect ADBC and TCNs to continue existing in the same form over the next 10 years (directed to NSF observers)?

- ADBC was developed to get digitization programs up and running and to give institutions resources to start digitizing their collections.
- NSF envisions the next 6 years will be mostly similar to what the program is doing now because there is still a long way to go.
    - The project only has around 500 collections (200 institutions) of the approximately 1500 collections in the U.S.
- NSF really wants this project to change attitudes in collection management and shape interest in digitization and data sharing.
- NSF anticipates that somewhere around year 8 there will be discussion about where the project should go next.
- NSF stressed that as a "national resource" where data are available to everyone, iDigBio could be a role model for other large projects.

Specific Advice

- iDigBio should incorporate the Smithsonian Institution's (SI) data.
    - SI would also like to participate more actively with the TCNs.
- iDigBio should determine what exactly its niche is.
    - Should iDigBio focus efforts in one realm, or plan to cover all of these topics but have overlap with multiple entities?
    - The key will be to determine iDigBio's unique contribution and then sustain only the unique infrastructure. iDigBio can then partner to cover gaps and/or overlap.
- iDigBio should seek to answer: Who are your key partners? Who are your customers? What do you need to be able to deliver? The answers to these questions will help guide what to focus on in years 6-10.
- There may be opportunities to develop monetary contracts with institutions to run digitization programs and supply data storage and personnel training. This could be one aspect of a long-term sustainability strategy.
- The research uses seems to be developing naturally, but iDigBio should increase engagement with all user groups which may benefit from specimen data.
- iDigBio serves as a network to bring together data from taxonomic experts and could also serve as a network for expert services and interpretations of these data. iDigBio should explore possible models to provide an integrated network for data and expertise, perhaps through partnership with expert societies.
- iDigBio should consolidate its successes and continue with the roles it has already undertaken (e.g., connecting the community, facilitating digitization, etc.). At the same time, iDigBio should ensure that its activities will support expected uses and that all data mobilized through TCNs are managed and curated in a persistent and sustainable way.
- A diversity of interest communities are mobilizing on the web and providing a lot of energy in amateur naturalist communities forming around taxonomic groups. iDigBio could use these communities as a

resource. These amateur communities can be more than customers; they could be partners (para-curators).