iDigBio Strategic Plan

OVERVIEW

Integrated Digitized Biocollections (iDigBio) is the central resource for the Advancing Digitization of Biodiversity Collections (ADBC) program funded by the U.S. National Science Foundation (NSF). Through ADBC, data and images for millions of biological specimens are being made available in electronic format for the research community, government agencies, students, educators, and the general public.

This Strategic Plan provides a roadmap and vision for the iDigBio project and its five domains: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure.

VISION

The vision for ADBC is a permanent database of digitized information from all biological collections in the United States that leads to new discoveries through research and a better understanding and appreciation of biodiversity through improved education and outreach, which then leads to improved environmental and economic policies. Creation of the digitized database is occurring in four stages:

1. An initial stage in which the effort to digitize biological collections across the U.S. is catalyzed by funding from NSF and from effective iDigBio-driven activities that foster collaborations, identify priorities, and generate information on best practices related to standards, workflows, and data management for digitization of biological collections, as well as demonstrate the value of biodiversity and collections that document biodiversity.
2. An intermediate stage wherein digitization at Thematic Collections Networks (TCNs), Partners to Existing Networks (PENs), and other participating institutions and networks improves methods and strategies and demonstrates the scientific and societal benefits of validated and readily accessible data.
3. A third stage in which the vision for ADBC is realized through participation by all U.S. institutions with biological collections.
4. A fourth stage in which digitization is a routine and sustained practice in all institutions with biological collections, and the national database is easily accessible as an up-to-date source of information on biodiversity.

MISSION

The mission of iDigBio is to develop a national infrastructure that supports the vision of ADBC by overseeing implementation of standards and best practices for digitization; building and deploying a customized cloud computing environment for collections; recruiting and training personnel, including underserved groups; engaging the research community, collections community, citizen scientists, and

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210)

the public through education and outreach activities; and planning for long-term sustainability of the national digitization effort.

HORIZON

iDigBio is enabling digitization of data from all U.S. biological collections and is integrating those data to make them broadly available and useful with shared standards and formats. Ultimately, ADBC is furthering the discovery and understanding of biological diversity, and iDigBio is engaging the research, collections, and education communities in a spirit of collaboration in an effort to open biological research collections to new downstream user communities.

iDigBio involves the development of a permanent and powerful cloud computing infrastructure to link biological data from collections across the U.S. into a single unified web interface, overcoming the "data silos" that currently exist across the country. Search and analytical tools enable users to mine diverse data, such as taxonomy, geographic location, 2- and 3-dimensional images, vocalizations, and molecular resources tied to specimens in collections. These data promote integrative biological research on living and fossil species and provide an immense resource for agricultural science and land use management, human health, and assessing the impacts of climate change, invasive species, and other natural resource management issues.

Key partners in this effort are the Thematic Collections Networks (TCNs), which form a national grid of institutions that are digitizing specimens and associated resources. Integration with the greater community of biocollections resources, tools and organizations is critical to accomplishing the grand challenge of digitizing and integrating data from all U.S. collections, large and small. For more details regarding the larger community that encapsulates ADBC, please refer to the Network Integrated Biocollections Alliance (NIBA) strategic plan and implementation plan.

PROJECT ORGANIZATION

ORGANIZATIONAL STRUCTURE

The iDigBio project is organized into five domains to achieve its goals: (1) Project Administration & Management, (2) Education & Outreach, (3) Serving the Research Community, (4) Digitization, and (5) Cyberinfrastructure. Leadership for each domain is provided by the Principal Investigators (PIs).

PROJECT OVERSIGHT

The project incorporates several oversight committees to provide governance, advice, and leadership over goals, strategies, implementation, activities, and progress. Project decisions impacting project scope, budget, or risk are made by the Project Director in consultation with the Executive and/or Steering Committee, while day-to-day planning and decision-making fall to the Project Manager.

Executive Committee

The Executive Committee (EC) is composed of the iDigBio Project Director, Principal Investigators, and Project Manager. The Executive Committee is responsible for overall project management, general oversight of iDigBio activities, managing conflicts of interest, and implementation of the strategic plan, including assurance that the digitization, research and educational missions of iDigBio are integrated and accomplished.

Steering Committee

The Steering Committee (SC) is composed of the iDigBio Project Director, Principal Investigators, Project Manager, and other Senior Personnel. The Steering Committee is responsible for reviewing progress and coordinating activities in digitization, research, education and outreach at iDigBio, and for advising others in iDigBio on overall resource allocation, strategic directions and management policies.

Internal Advisory Committee

The Internal Advisory Committee (IAC) is composed of the Project Manager, Biodiversity Informatics Manager, and representatives from the Thematic Collections Networks (TCNs) and other digitization projects and collections working with iDigBio. The IAC meets regularly to report on progress in digitization efforts, share and identify best practices and standards, identify gaps in digitization areas and technology, and enhance training efforts.

External Advisory Board

The External Advisory Board (EAB) is composed of members selected by iDigBio and approved by the NSF program officer. The EAB meets annually and is responsible for advising iDigBio on its strategic directions, management policies, and activities, including progress and integration of digitization projects, research, education and outreach activities among all funded institutions.

STRATEGIES AND OBJECTIVES

PROJECT ADMINISTRATION & MANAGEMENT

iDigBio Project Administration & Management decisions and activities are executed under the leadership of Dr. Larry Page, iDigBio Project Director, who is based at the University of Florida's Florida Museum of Natural History. The Project Administration & Management domain provides administrative and logistical support to the other four domains and ensured that activities are coordinated within and across iDigBio domains and with collaborative organizations as appropriate. The key objectives of the Project Administration & Management domain are:

- Coordinate project oversight committees and other key meetings
- Plan for long-term sustainability, including participation with NIBA

- Comply with NSF cooperative agreement and reporting requirements
- Build and maintain strategic partnerships
- Plan and monitor the project budget
- Manage internal and external communications
- Develop and maintain collaboration and communication capabilities
- Serve as the focal point for ADBC to help resolve TCN problems, remove roadblocks, and answer digitization-related questions from the community

Education and Outreach

iDigBio Education and Outreach (E&O) decisions and activities are executed under the leadership of Dr. Bruce MacFadden, iDigBio co-PI, who is based at who is based at the University of Florida's Florida Museum of Natural History. iDigBio E&O activities are focused on digitization curricula development, stakeholder identification, E&O materials and protocols, and public speaking engagements, which are recorded and published for national impact. The key objectives of the iDigBio E&O domain are:

- Foster project awareness within the scientific and collections communities
- Engage the public collections community through resources, deliverables, and opportunities that highlight the importance of biodiversity collections and digitization
- Identify and assess the needs of target audiences, downstream user groups, and other stakeholders
- Identify downstream partners and stakeholders, document their needs, and execute outreach activities to meet those needs
- Develop educational resources related to digitization and biodiversity
- Measure the impact and success of intended learning outcomes

Serving the Research Community

iDigBio decisions and activities related to effectively serving the research community are executed under the leadership of Dr. Pamela Soltis, iDigBio co-PI, who is based at who is based at the University of Florida's Florida Museum of Natural History. The Research domain leverages existing relationships to deliver branding and messaging related to iDigBio in efforts to foster collections/research community adoption of iDigBio's services, infrastructure, tools, resources, and data: specimen portal, website, Wiki, Forums, Listservs, Adobe Connect teleconferencing, Workshop coordination/funding, Working Group coordination/funding, Appliance development, and Hosting of services. In addition, the Research domain highlights developments and opportunities for research using specimen data alone or in conjunction with other types of data. The key objectives of the Research domain are:

- Engage the research community to promote community adoption of iDigBio's services, infrastructure, tools, resources, and data
- Produce detailed use cases of research applications of specimen data, provide them to the Cyberinfrastructure team, and help validate the effectiveness of implementation
- Seek opportunities for integration of iDigBio specimen data and API services with key data and research services from other projects and organizations
- Identify strategic partners in the research and collections community and develop synergistic relationships with those partners

DIGITIZATION

iDigBio decisions and activities related to optimization of digitization workflows and processes, digitization documentation, and efforts to share and improve digitization tools within the collections community are executed under the leadership of Dr. Greg Riccardi, iDigBio co-PI, who is based at Florida State University's Institute for Digital Information and Scientific Communication. iDigBio digitization experts work with the community to understand gaps in workflows, processes, practices, and tools that prevent effective and efficient specimen digitization. These activities are catalyzed by site visits, virtual and on-site workshops, working groups, contact with tool developers, and user contact to foster a high-degree of community involvement. The observations and subsequent analyses from these activities yield documentation, papers, training materials, and presentations that promote effective digitization practices and workflows, including qualitative and quantitative measures of success. The key objectives of the Digitization domain are:

- Engage the collections community to market and build interest in adopting iDigBio's services, infrastructure, tools, resources, and data
- Establish minimum information standards and data fitness for use parameters
- Optimize digitization workflows
- Conduct digitization training and produce online training materials
- Enhance and broaden exposure to digitization tools and resources including Georeferencing, Optical Character Recognition (OCR), Natural Language Processing (NLP), Authority Files, Optimized Digitization Workflows, and Crowdsourcing
- Evaluate, document, and publish analyses related to digitization hardware and software tools
- Identify significant technological gaps in digitization capabilities that require additional resource investment in order to ensure the success of ADBC

CYBERINFRASTRUCTURE

iDigBio Cyberinfrastructure decisions and activities are executed under the leadership of Dr. José Fortes, iDigBio co-PI, who is based at the University of Florida's Advanced Computing and Information Systems (ACIS) laboratory. Cyberinfrastructure decisions are informed by the experience and expertise of ACIS

personnel, key stakeholders in iDigBio, biodiversity community input, and information technology community input. The Cyberinfrastructure implementation process follows a formula that balances strategic planning with the agility to meet new challenges, short-term project needs, and enhanced/clarified specifications. The cyberinfrastructure team releases new features and upgrades to the specimen portal on a semi-annual basis. The key objectives of the Cyberinfrastructure domain are:

- Research, implement, and maintain a scalable cloud infrastructure for text (data/metadata) and object (media) storage
- Implement infrastructure to enable hosting of web services and/or websites for strategic partners, such as TCNs
- Deploy iDigBio appliances and services via multiple channels (e.g., web services, locally-run virtual machines, infrastructure-as-a-service cloud implementations) to enhance, simplify, and/or improve activities of data providers and data consumers
- Implement a comprehensive authentication and access control system to enable data tracking and a cohesive user experience among iDigBio's systems
- Develop, implement, and maintain iDigBio APIs to access text and media data stores
- Develop, implement, and maintain a Graphical User Interface (GUI) to provide end-users, including data contributors and data consumers, access to search, visualize, and download text and media data from the cloud infrastructure
- Integrate iDigBio services and user portals with key strategic partners and other collaborators
- Secure infrastructure resources to maintain adequate performance and capacity

MEASUREMENT OF SUCCESS

iDigBio has a Project Evaluator on staff, a professional with expertise and training in evaluation, who carries out formal evaluations of iDigBio-supported activities. These evaluations are made available to organizers of these activities, with the objective of providing guidance on best practices for future activities. iDigBio also utilizes evaluation results to assess effective interdisciplinary scientific collaborations, including the potential for new media or formats to aid collaborative efforts. Overall evaluation of iDigBio involves metrics including: new collaborations initiated, data ingested into the specimen portal, products derived from iDigBio-supported activities, education of individuals within the community, and contributions towards the general understanding of the value of digitization of biodiversity collections.

ENVIRONMENTAL SCAN

The following is a summary of the major strengths, weaknesses, opportunities, and threats that have been highlighted by the External Advisory Board, NSF Site Review team, and surveys of the collections community.

	POSITIVE	NEGATIVE
INTERNAL	**Strengths – iDigBio has…** • Successfully engaged with the collections community, particularly through workshops and working groups • Developed a vast repository of resources information • Optimized digitization protocols and workflows • Actively reduced barriers to digitization • Demonstrated strong leadership and management • Developed partnerships and collaborations, including significant synergy with TCNs, networking, and building community	**Weaknesses – iDigBio has…** • Non-intuitive website navigation and inconsistent arrangement of resources • Minimal data in specimen portal • Inconsistent internal communication • Insufficient involvement of smaller collections and institutions • Insufficient effort to promote diversity and include under-represented groups • Lack of explicitly defined of project leadership evaluation plans
EXTERNAL	**Opportunities – iDigBio should…** • Highlight research uses of data • Develop digitization standards and encourage compliance • Advocate for greater digitization funding and formation of cooperative groups across institutions and projects • Coordinate efforts across institutions, disciplines, and agencies to bridge initiatives and facilitate networking • Integrate directly with collection management software and tools • Capture “dark data” from both large and small institutions • Collaborate with key existing projects, such as Global Biodiversity Information Facility (GBIF) and Biodiversity Information Serving Our Nation (BISON)	**Threats – iDigBio should…** • Develop a strategy for financial sustainability to reduce dependence on federal funding • Develop a strategy for long-term storage and maintenance of its cyberinfrastructure, metadata and images • Be mindful of its extremely broad project scope, which includes a vast array of collaborations and partnerships • Be mindful that iDigBio depends on a number of external projects, tools, and initiatives whose long-term sustainability could be questionable

SUSTAINABILITY PLAN

To ensure long-term sustainability of the national digitization effort, iDigBio is implementing strategies to maintain its strengths, address its weaknesses, capitalize on available opportunities, and mitigate potential threats.

MAINTAINING STRENGTHS

The Integrated Digitized Biocollections (iDigBio) project has made significant progress since the initiation of funding in 2011. iDigBio's innovations include both sociological and technological accomplishments with wide-ranging benefits to the collections community.

First and foremost, iDigBio has established successful ***communication*** between the Information Technology (IT) and biological collections communities. Having bridged this "cultural" barrier, iDigBio personnel are working together to identify challenges and to design appropriate solutions. This communication extends beyond the personnel specifically working on iDigBio to other partners, such as the TCNs, which allows for collaboration, synergy, and effective training throughout the community.

Perhaps the most successful innovation of iDigBio to date is the series of ***training workshops*** that have been organized and sponsored by iDigBio personnel. These workshops have delivered effective training on digitization-related methods and practices, as well as on other topics contributed by the workshop participants. These workshops and training materials, publically available at www.idigbio.org, have provided a wealth of new resources and have secured iDigBio's leadership role in ***workforce development*** within the collections community. During its first 3 years, iDigBio has sponsored the attendance of over 1265 participants (over 700 of which are unique) from over 300 unique institutions to its 38 workshops, summits, symposia, and other events.

Finally, the forthcoming ***availability of massive amounts of specimen data*** has energized the collections community about the use of specimen data for a variety of big research questions that have been intractable to this point. This renewed energy within the community has fostered iDigBio to produce significant innovations in IT design and implementation, including: Creating the practice of introducing identifiers in the data stream to enable data linking; Development of emerging data models for ingestion and integration of data sets from diverse collections; Cloud architecture for data storage, retrieval, and management; an Open schema infrastructure to offer flexibility and agility in handling an evolving data model; and an Appliance framework to respond to the needs of biocollections informatics.

ADDRESSING WEAKNESSES

iDigBio has recognized that its website and specimen portal are the primary interfaces between the project and its audiences. Completely revised versions of the iDigBio website and specimen portal were released to the community in December 2013. The website was redesigned to focus on making it easier

to understand and to use and to be more approachable to a lay visitor. The specimen portal and APIs were redesigned to create the foundation for a system that will serve the community for years to come, including user interface improvements and improved stability and flexibility.

iDigBio has further recognized that the availability of specimen data in the portal is highly important. As such, iDigBio has moved data ingestion to the forefront and has worked diligently to improve the efficiency of the data mobilization and ingestion process. iDigBio has the goal of ingesting data sets within 2 weeks of data mobilization completion. Data mobilization is dependent on the provider, although iDigBio staff continually track the process from initiation to closure using Redmine. From December 2013 thru April 2014, the number of collections in the portal has increased from 121 to 235 (↑94%), the number of specimen records has increased from 4,410,237 to 14,036,195 (↑218%), and the number of media records has increased from 1,005,679 to 2,290,983 (↑128%).

iDigBio senior personnel have acknowledged that communication, both within iDigBio and with our clients, could be improved. iDigBio believes that less-than-ideal communication is the result of having a small staff charged with multiple tasks who are unable to find the time to report on activities in a meaningful way and in a reasonable timeframe. In response, iDigBio has hired a Communications Coordinator who is charged with tracking iDigBio activities and broadly disseminating information about those activities. The Communications Coordinator has begun the task of improving communications by defining categories of participating audiences and then focusing high-impact messages tailored for those audiences using a variety of media outlets. iDigBio personnel are making a concerted effort to write and post concise reports on our website in a timely fashion, and the Communication Coordinator announces their release using Twitter, Facebook, and other more traditional means.

iDigBio has initiated efforts to provide national leadership in developing methods to enhance the cultural and ethnic diversity of participation in the biological sciences. iDigBio is hosting a series of workshops targeted at undergraduate students in underrepresented populations focusing on opportunities for careers and graduate study in environmental biology, biodiversity, ecology, and evolution.

iDigBio has initiated efforts to annually evaluate the project and project leadership. The project is evaluated by the collections community to gauge interest and involvement with iDigBio and to solicit input on how iDigBio is doing. The project leadership is evaluated by iDigBio staff and students, giving them an opportunity to share their views regarding their own roles and responsibilities and to solicit input on project leadership and management.

Capitalizing on Opportunities

iDigBio is working to highlight research uses of the data in the specimen portal. iDigBio has begun implementing features in the specimen portal that are based directly upon input from biologists and

how they anticipate using the data provided. For example, iDigBio is working to implement advanced mapping features in the specimen portal. In addition, iDigBio is working with GenBank to create links between specimen data and genetic data.

iDigBio is continuing to advocate for, coordinate with, and facilitate networking among the collections community through its involvement in professional meetings and societies. iDigBio is also working to secure direct representation in key organizations, such as The Society for the Preservation of Natural History Collections (SPNHC), Natural Science Collections Alliance (NSCA), and the American Institute of Biological Sciences (AIBS).

iDigBio is continuing to support a variety of working groups and interest groups (https://www.idigbio.org/wiki/index.php/IDigBio_Working_Groups). iDigBio is allowing these groups to form organically to dynamically address the needs of the collections community. Some of iDigBio's groups are focused on delivery of short-term objectives, while others are tasked with ongoing research, development, and improvement activities.

iDigBio is continuing to develop its partnerships with Specify, Symbiota, and KE EMu to introduce globally unique identifiers and to create export features specifically tailored to iDigBio. The goal of these partnerships is to streamline the path to data ingestion with iDigBio.

iDigBio has started to mobilize data sets beyond the TCNs by using the Global Registry of Biorepositories (GRBio). iDigBio has developed a partnership with GRBio and is making contact with U.S. institutions outside the TCNs. In addition, iDigBio is advancing its existing partnerships with the Global Biodiversity Information Facility (GBIF) and Biodiversity Information Serving Our Nation (BISON).

Mitigating Threats

iDigBio, in concert with the Network Integrated Biological Alliance (NIBA), is examining and developing business models that create an environment for long-term economic sustainability. Examples of such business models include cost-recovery, endowments through benefactors, and engaging in public/private partnerships. The goal is to establish a multifaceted funding base for long-term sustainability through community engagement and partnerships with government agencies, private foundations, and international stakeholders.

Building a community of use around iDigBio is paramount to sustaining the infrastructure. As the success of ADBC continues, collections-based data are becoming much more generally available, used, and appreciated, not only for their scientific value, but also for understanding the importance of biological diversity to the economic stability and health of the planet. Coupled with tools from informatics and computer science, collections data are being made available for a host of applications related to biodiversity research, natural resource management, and public policy development. iDigBio has

developed a data mobilization and ingestion process to ensure the vast number of collections being harvested are tracked from initiation to closure using Redmine.

iDigBio remains vigilant in managing its scope. To reduce uncertainty and to prevent scope creep, iDigBio has published the Project Scope along with a series of agreed upon in-scope and out-of-scope activities on the iDigBio website. The iDigBio Steering and Executive Committees regularly review tasks, objectives, and progress to insure they are within scope.

iDigBio is establishing strategic partnerships with the many external projects, tools, and initiatives that it depends upon. These strategic partnerships not only promote integration of tools and services into iDigBio but also promote long-term sustainability of both iDigBio and its dependencies. Ultimately, these strategic partnerships will contribute towards a robust comprehensive infrastructure.