IMPLEMENTATION PLAN FOR THE NETWORK INTEGRATED BIOCOLLECTIONS ALLIANCE

IMPLEMENTATION PLAN FOR THE NETWORK INTEGRATED BIOCOLLECTIONS ALLIANCE

FINAL REPORT OF A WORKSHOP*
ORGANIZED BY THE

AMERICAN INSTITUTE OF BIOLOGICAL SCIENCES

WITH SUPPORT FROM THE
NATIONAL SCIENCE FOUNDATION
Directorate for Biological Sciences
Division of Emerging Frontiers

September 17–18, 2012
Herndon, VA

* Participants (boldface denotes writing committee): C Bartlett, **J Beach**, N Cobb, **J Cook**, C Dietrich, **LS Ford**, J Fortes, S Graves, C Gries, **R Gropp**, G Guala, **J Hanken (co-organizer)**, **K Joyce (facilitator)**, MA Mares, R McCourt, **L McDade (co-organizer)**, A Neill, C Norris, L Page, C Parr, G Riccardi, N Rios, K Seltmann, D Smith, **B Thiers**, and Q Wheeler.

This material is based upon work supported by the National Science Foundation under Grant No. EF-1241179.

This document is available from the American Institute of Biological Sciences, 1900 Campus Commons Drive, Suite 200, Reston, VA 20191 or online at *www.aibs.org/public-policy/biocollections.html*.

Cover images: Florida Museum of Natural History, Agricultural Research Service, Graham Snodgrass, and Julie Palakovich Carr.

Implementation Plan for the Network Integrated Biocollections Alliance

Executive Summary

The biological collections (biocollections) in the United States are the result of nearly 250 years of scientific investigation, discovery, and inventory of living and fossil species from this country and around the world. Scientists have amassed, annotated, and curated more than one billion specimens in the more than 1600 biocollection institutions across the United States. These specimens and their associated data are maintained for research and education and to inform wise decisions about the environment, public health, food security, and commerce.

This monumental investment of human capital and financial resources in species discovery, documentation, and analysis is active and ongoing. Moreover, the specimens and data in biocollections are of value to many people beyond biologists. Computer scientists, geologists, informaticists, environmental scientists, land managers, educators, and citizen scientists are among the communities increasingly seeking access to this vital resource.

Recognizing the significant value of biocollections for research, education, and society, the biocollections community coalesced in 2010 to develop *A Strategic Plan for Establishing a Network Integrated Biocollections Alliance* (NIBA). The plan outlines the elements required for an "inclusive, vibrant, partnership of US biological collections that collectively will document the nation's biodiversity resources and create a dynamic electronic resource that will serve the country's needs in answering critical questions about the environment, human health, biosecurity, commerce, and the biological sciences." The plan issues a strong and urgent call for an aggressive, coordinated, large-scale, and sustained effort to digitize the nation's biological collections in order to mobilize their data (including images) through the Internet.

Federal agencies and the scientific community have begun to respond to the NIBA strategic plan. The biocollections community now recognizes, however, the need for an implementation plan that explicitly identifies the corresponding actions, timelines, and milestones required to achieve the goals of the *Strategic Plan*. In September 2012, the American Institute of Biological Sciences (AIBS), with support from the US National Science Foundation (NSF), convened a workshop of experts in biocollections, digitization, computer science, and other relevant fields to develop this *Implementation Plan* for NIBA.

The NIBA implementation plan provides a solid, realistic, and effective framework for achieving the three key objectives articulated in the NIBA strategic plan: (1) to "digitize data from all US biological collections, large and small, and integrate these in a Web-accessible interface using shared standards and formats"; (2) to "develop new Web interfaces, visualization and analysis tools, data mining, georeferencing processes and make all available for using and improving NIBA resources"; and (3) to "create real-time upgrades of biological data and prevent the future occurrence of non-accessible collection data through the use of tools, training, and infrastructure."

This *Implementation Plan* has been informed by other international, national, and regional scientific and technical initiatives and activities; by participants in the September 2012 workshop; and by comments solicited from current and potential stakeholders. It provides detailed recommendations towards achieving the following goals: (1) to establish an organizational and governance structure that will provide the national leadership and decision-making mechanism required to implement NIBA and to fully realize its *Strategic Plan*, (2) to advance engineering of the US biocollections cyberinfrastructure, (3) to enhance the training of existing collections staff and to create the next generation of biodiversity information managers, (4) to increase support for and participation in NIBA by the research community and a broad spectrum of stakeholders, (5) to establish an enduring and sustainable knowledge base, and (6) to infuse specimen-based learning and exploration into formal and informal education.

This plan reflects a coordinated and collaborative effort to realize the grand goals proposed by the biocollections community in the NIBA strategic plan. All are welcome and encouraged to participate.

Background

More than 1600 collection institutions distributed across the United States curate vast numbers of biological specimens and their associated data. During nearly 250 years of survey and inventory, United States–based scientists have amassed and studied more than a billion such specimens, including fossils. Scientists use these specimens daily in research and education to identify and classify species, to document the dynamic distribution of life on Earth, and to provide knowledge to inform our understanding of evolutionary and environmental change and threats to public health. This is a truly monumental investment of human capital and financial resources in species discovery, documentation, and analysis, and it is active and ongoing. The accumulated knowledge regarding the world's species that is contained in these collections makes them national treasures of immense value for future biological research; education; and informed environmental, agricultural, and public health policy.

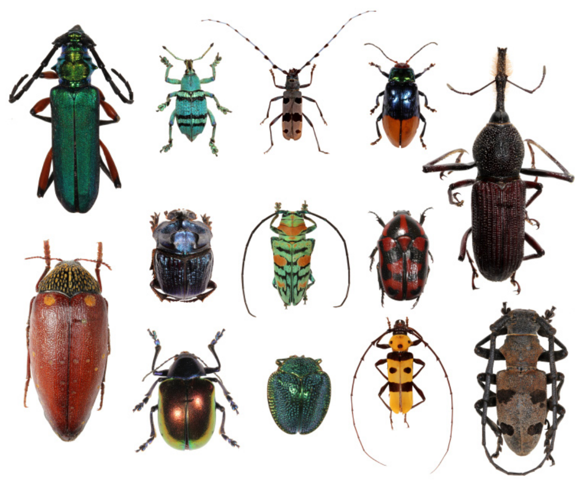
Specimens and data in biocollections are of value to people far beyond the community of biodiversity scientists. Computer scientists, geologists, informaticists, environmental scientists, and many others are increasingly seeking access to biocollections as part of their research. In addition to informing responses to society's greatest environmental challenges and enhancing our understanding of Earth's natural history, biocollections are stimulating the development of innovative technology, informing the creation of new data-use and management tools, engaging citizen scientists, and enhancing both formal and informal science education.

The NSF has been the essential catalyst for improving the accessibility and use of our nation's biocollections that are held and cared for by organizations outside the federal government. The NSF has supported programs that stimulate collections-based research in systematic biology and biodiversity inventories and the development of new research instrumentation and technologies. As a result of these and other government-sponsored and community-sourced initiatives, as early as the 1960s, biodiversity researchers began proposing large, nationally coordinated efforts that would use emerging computer, database, and imaging technologies to advance biodiversity research. Internet-accessible, aggregated (or federated) specimen databases organized around major taxonomic groups were subsequently created and brought online. These include FishGopher (1993), FishNet (1999), MaNIS (2001), HerpNET (2002), ORNIS (2004), FishNet2 (2005), and VertNet (2012).


Florida Museum of Natural History

Over the last two decades, several other countries have launched programs to digitize data from specimens held in their biocollections and make these data readily accessible online. Notable examples are Australia (ALA), Costa Rica (INBio), and Mexico (CONABIO). Equally important, corresponding efforts to achieve international integration and interoperability of these and other data were also launched, including the Global Biodiversity Information Facility (GBIF) and the Global Plants Initiative (GPI).

These initiatives have helped both scientists and policymakers recognize the scientific and societal benefits that emerge when researchers, educators, and other stakeholders have the ability to efficiently access information from and about biocollections. Therefore, in 2005, President George W. Bush chartered the federal Interagency Working Group on Scientific Collections (IWGSC) to evaluate and make recommendations about the needs of the federal government's scientific collections. President Barack Obama's administration has furthered this effort, including a directive issued by the Office of Science and Technology Policy (OSTP) that requires federal agencies to budget for science collections. Through §104 of the America COMPETES Act Reauthorization of 2010 (P.L. 111-358), Congress directed OSTP to "improve the quality, organization, access, including online access, and long-term preservation of such collections for the benefit of the scientific enterprise," and to work with representatives of nongovernmental organizations and institutions that have a "stake in the preservation, maintenance, and accessibility of such collections."



In addition to participating in the IWGSC, the NSF has taken a leadership role in gathering data about nonfederal science collections [1] and supporting Research Coordination Networks (RCNs) and other initiatives that strengthen the biocollections community. These data and several related workshops sponsored by the NSF have identified common needs, including enhanced coordination and networking among collections and curators, finding and retaining expertise, reducing the risk of specimen loss, improving the accessibility of collections, and developing new tools to enable the exchange of digital data.

In 2010, the biocollections community coalesced around a plan that addresses these needs. The NIBA strategic plan outlines the elements required for an "inclusive, vibrant, partnership of US biological collections that collectively will document the nation's biodiversity resources and create a dynamic electronic resource that will serve the country's needs in answering critical questions about the environment, human health, biosecurity, commerce, and the biological sciences." The *Strategic Plan for Establishing a NIBA* issues a strong and urgent call for an aggressive, coordinated, large-scale and sustained effort to digitize the nation's biological collections in order to mobilize their data (including images) through the Internet. It also is closely aligned with the NSF's current and broad-based Big Data initiative.

In response to recommendations in the NIBA strategic plan, in August 2010, the NSF established the Advancing Digitization of Biodiversity Collections (ADBC) program. Through ADBC, the NSF is providing seed money to initiate a 10-year effort to fully digitize United States–based collections. The NSF has pledged to provide at least $100 million over this period, or about $10 million per year. The ADBC program seeks to digitize specimens efficiently and cost effectively while focusing on specimens needed to answer grand research challenges for biodiversity, to fund collaborative Thematic Collections Networks (TCNs) based on major research challenges, and to fund a central resource or hub to help support the work of the TCNs. Importantly, NSF funding supports digitization and research programs only at nonfederal collections; it may not be used to digitize specimens and specimen-associated data owned by the federal government. Support for digitization efforts that target the many hundreds of federal collections falls to the agencies responsible for these specimens and data. Continued and increased collaboration among these agencies and between these agencies and nonfederal collections will promote wise resource allocation and the maximum use of biocollections for research, education, and decision making.

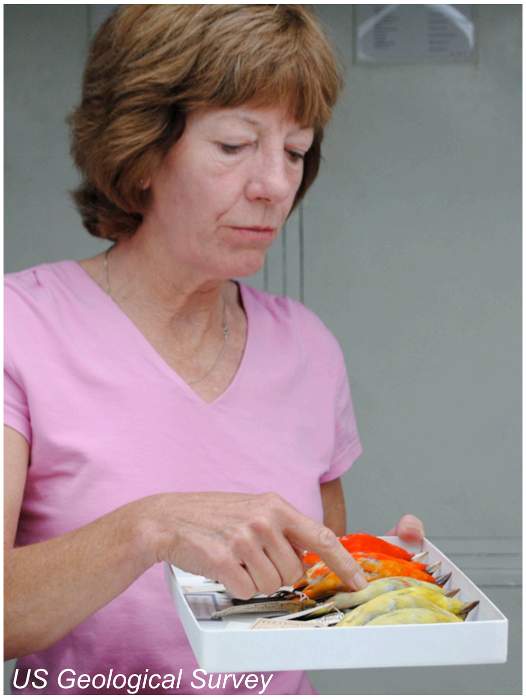
Thus far, the NSF has funded a central resource, iDigBio (Integrated Digitized Biocollections, based at the University of Florida and Florida State University), which serves as the coordinating center for the national digitization effort. iDigBio supports the digitization of data by overseeing the implementation of accepted standards and practices; promotes data integration and connectivity to make them broadly available to the scientific community and other stakeholders through a common portal and underlying cyberinfrastructure; and facilitates research, educational, and other outreach activities.

The NSF has also funded seven TCNs. These networks involve 130 institutions in 48 states; they are intended to digitize and unite data from approximately 65 million specimens. The technical issues addressed by these groups include the integration of heterogeneous data, challenges associated with digitizing insect collections, how to use citizen scientists to increase the speed of the digitization process, how to integrate ancillary material into the digitization process, and how to improve technology to expedite imaging. Current TCN grants are addressing the following topics: *North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change*; *Plants, Herbivores and Parasitoids: A Model System for the Study of Tri-Trophic Associations*; *InvertNet—An Integrative Platform for Research on Environmental Change, Species Discovery and Identification*; *Mobilizing New England Vascular Plant Specimen Data to Track Environmental Changes*; *Digitizing Fossils to Enable New Syntheses in Biogeography—Creating a PALEONICHES-TCN*; *The Macrofungi Collection Consortium: Unlocking a Biodiversity Resource for Understanding Biotic Interactions, Nutrient Cycling, and Human Affairs*; and *Southwest Collections of Arthropods Network (SCAN): A Model for Collections Digitization to Promote Taxonomic and Ecological Research*.


Florida Museum of Natural History

In parallel with the above initiatives and achievements, the NSF also funded *Building a National Community of Natural History Collections*, an RCN that seeks to establish natural history collection goals for the twenty-first century and to address them with a unified voice. Partners in this effort included AIBS, the Natural Science Collections Alliance (NSCA), and the Society for the Preservation of Natural History Collections (SPNHC). The RCN sponsored several successful workshops, symposia, and internships. Although digitization was not a central focus, related themes emerged that affect digitization efforts. These include (1) the need to develop new technologies to digitize collections and facilitate taxonomy while upgrading the physical infrastructure of collections, (2) a strong call to unify efforts among collections on common issues and to increase student and public involvement in collection efforts, (3) recognition that small collections have the potential to innovate tools and techniques for digitization, and (4) how the aggregation of mass data can both increase research opportunities across historical taxonomic barriers and revolutionize educational opportunities and public outreach.

At about the same time, the NSF also funded a series of workshops on the *Future of Systematics and Biodiversity Science*. Led by Patrick Herendeen, Lucinda McDade, and Petra Sierwald, four workshops, involving more than 100 participants, were convened between May 2009 and September 2010. Although the workshops were not directly tasked with considering biocollections-based issues, a number of the resulting initiatives and products are relevant. The participant group as a whole formally endorsed the NIBA strategic plan in a letter submitted during the public review phase. In a commentary authored by a subset of workshop participants, McDade and colleagues [2] identified several problems associated with the professional review and advancement of scientists who devote considerable effort to the curation of biological collections; these contributions are undervalued in comparison to grants and publications, yet they are vital to the NIBA objectives.


US Geological Survey

In 2010, the NSF's Office of Cyberinfrastructure funded two workshops as part of its Scientific Software Innovation Institutes program. In these workshops, held in 2011 and 2012, the participants analyzed technology and workflow-design options for biological specimen digitization.[3] The national partners in the program included SPNHC, the Encyclopedia of Life's Biodiversity Synthesis Center, and iDigBio. The participants in both workshops concluded that the most significant challenge facing a national biocollections digitization initiative is the design and support of extensible software, protocols, and community standards needed to streamline the process of capturing and mobilizing collections data. The corresponding cyberinfrastructure will need to accommodate diverse types of specimen preparations and discipline-specific curatorial protocols yet must be fully capable of integrating the output of local and project-specific workflows into a national resource. The workshop participants also proposed the creation of a software-engineering institute for biodiversity informatics, which would have three principal roles: (1) to develop and support new digitization technologies that emerge from NSF ADBC–funded TCNs; (2) to acquire and adapt technological advances related to specimen digitization from engineering, computer science, and library and information science; and (3) to collaborate with iDigBio to provide help-desk support and dissemination of new software tools to the collections community.

The recently funded AIM-UP! (Advancing Integration of Museums into Undergraduate Programs) is an RCN–UBE (Undergraduate Biology Education) that explores novel ways to incorporate the vast bioinformatics resources of museums into problem-based lessons for college undergraduates. These archives and databases have been the basis for renowned research and graduate-training programs, but they are generally less well integrated into undergraduate education.

In addition to work supported by the NSF, the US Geological Survey (USGS) is developing a program—Biodiversity Information Serving Our Nation (BISON)—that will contribute significantly to the implementation of NIBA. It is anticipated that BISON will perform at least two important functions. First, it will serve as an integrated resource for biological occurrence data from the United States and will function as the US node in GBIF. Second, it will link these data to emerging EcoINFORMA [4] activities in the federal government and will thereby mobilize and integrate environmental data for sustaining the nation's environmental capital.

Despite these initial efforts by the NSF and USGS, many aspects of NIBA's strategic plan remain to be implemented. These include several concrete activities that must be completed if NIBA is to achieve its ambitious goal of an "inclusive, vibrant, partnership of US biological collections that collectively document the nation's biodiversity resources and create a dynamic electronic resource that will serve the country's needs in answering critical questions about the environment, human health, biosecurity, commerce, and the biological sciences." All relevant stakeholders (e.g., the scientific community, government agencies, data and knowledge managers and users, tool and technology developers, collection professionals, and institutions) must contribute and work together to implement NIBA.

On 17 and 18 September 2012, AIBS, with support from the NSF, convened a workshop in Herndon, Virginia. The workshop brought together a broad cross-section of experts to consider and recommend the best course for fully implementing NIBA (Appendix A). The participants included several individuals who had participated in earlier workshops, including those described above, that had led to the development of NIBA's strategic plan. Following the September workshop, a subset of eight participants constituted a writing committee, which drafted the present document. All workshop participants then reviewed and provided comments on the draft plan. Additional comments on the draft document were then solicited broadly across researcher and collections based areas (Appendix B) and used to prepare the final *Implementation Plan*.

Key Objectives from the NIBA Strategic Plan

"Digitize data from all US biological collections, large and small, and integrate these in a Web-accessible interface using shared standards and formats."

Achieving this objective will require consensus on the standards for definition and communication of specimen data objects for information exchange among institutional and project databases and the national aggregated data resource, as well as the provision of all specimen data through Web (HTML) and standard network application program interfaces (APIs).

"Develop new Web interfaces, visualization and analysis tools, data mining, georeferencing processes and make all available for using and improving NIBA resources."

Develop integrative software platforms for basic research, educational, and applied uses of specimen data in the national database. Incorporate or link to appropriate software applications when they already exist and encourage and enable the broader community to develop its own applications that both use and expand on existing platforms.

"Create real-time upgrades of biological data and prevent the future occurrence of non-accessible collection data through the use of tools, training, and infrastructure."

Derive novel data workflows that efficiently ingest information from historical specimens and promptly mobilize data from new specimens and species discoveries into the national data specimen resource.

Accomplishing these strategic plan objectives requires the completion of the biodiversity informatics core architecture, additional workforce training, increased educational engagement, new government and private sector partnerships, and a viable economic model for sustainability. This document lays out a corresponding implementation strategy that comprises six goals. Each goal, in turn, is associated with a series of implementation objectives ("Elements" in the NIBA strategic plan). We summarize the implementation objectives in table 1, which lists representative action items, their proposed timing and sequence, and current status. Together with other critical programs needed to support the development, maintenance, and use of our nation's biocollections, both federal and nonfederal, NIBA will deploy an essential resource that will serve our nation's government and citizens and help solve many of our most pressing environmental, societal, and scientific problems.

Goals

Goal 1: Establish an organizational and governance structure that will provide the national leadership and decision-making mechanism required to implement NIBA and to fully realize its strategic plan.

The Importance of Goal 1 for NIBA

Digitizing United States–based biological collections is a collaboration of grand scale that requires meaningful participation from all corresponding host institutions and funding sources. The input of the latter can help shape this initiative so that it is maximally effective; collections institutions and funding sources can also serve as ambassadors from the digitization effort to the broader communities that NIBA will ultimately serve. Yet, at this time, there is no organizational structure, formal or informal, that offers a unified framework for the community of collections professionals to help guide, facilitate, coordinate, and sustain all current digitization efforts. There is also no effective mechanism to advocate for and coordinate funding for collections digitization, especially one that addresses both federal and nonfederal collections. NIBA cannot succeed without effective organization and governance that provide leadership and decision making on a national scale, that are responsive to the needs of both the collections community and the likely consumers of digitized data and images, and that promote NIBA and advocate for its support at all levels.

The implementation objectives for this goal encompass a series of activities that, together, will achieve the required organizational and governance structure and will define a scope of work that is required to successfully manage and guide NIBA.

The Implementation Objectives for Goal 1

1.1. Establish the NIBA organization. Contract with an existing management organization, professional society, or other suitable entity to organize and lead NIBA for an initial period of 1–2 years while a more permanent management structure is determined and established. Among the first activities will be to convene one or two workshops that bring together representatives of NIBA stakeholder groups (e.g., collections-holding institutions, professional societies, current digitization initiatives, government agencies, knowledge consumers) to identify, evaluate, and recommend the most appropriate governance model or models to manage NIBA and achieve its implementation objectives in a timely and sustainable fashion. These workshops could produce recommendations for levels and sources of funding for NIBA, potential staffing requirements, methods for ensuring stakeholder engagement, and mechanisms to evaluate and ensure progress toward NIBA objectives.

1.2. Establish a governance, administrative, and management structure for NIBA. The governance body selected to implement NIBA would be responsible for both leading and coordinating the effort and for ensuring that it conducts the specific activities and achieves the ultimate goals set forth in the *Implementation Plan*. Configuring the governance model in this way will also help achieve appropriate and requisite buy in from stakeholders as well as address key issues, such as data stewardship and provenance, that are necessary for the initiative to succeed.

1.3. Establish metrics and mechanisms for measuring progress made toward NIBA's specimen digitization goals. The intent is to track progress toward agreed-on benchmarks and milestones for a given project, agency, or institution, as well as to evaluate the cost effectiveness of alternate methods of digitization or of the same method deployed at different institutions using different workflows or tools. Understanding the true cost of digitizing specimens and maintaining digital collections will be a critical, ongoing activity.

1.4. Document workflow challenges and seek solutions for "difficult" collections with special physical properties or handling needs. Convene a working group to identify those disciplines and collection sources that are the most difficult in regard to physical access to biodiversity data and their translation to a searchable format and propose realistic strategies to overcome these difficulties. The strategy should, in particular, address the needs of small collections, which may lack expertise in both digitization and data management (as well as core curatorial expertise). In developing its recommendations, the working group should evaluate digitization efforts pursued by other countries, which may offer effective and efficient approaches to be emulated by NIBA (e.g., Jisc's GB/3D Fossil Types online project). The working group should also review and build on the findings and recommendations offered in reports from prior workshops and collaborate with ongoing efforts (e.g., iDigBio).

1.5. Organize and facilitate both public and professional outreach to communicate NIBA's goals, programs, and accomplishments, as well as its broader societal benefits. These activities would heighten awareness and appreciation of biocollections and their relevance to contemporary science and society and would promote the use of digitized collections data and images in education, research, agriculture, forensics, human health, public policy, and other areas. It could include communications training for scientists to better enable them to convey technical information to policymakers, administrators, and nonprofessionals in an accessible and understandable form. Facilitating the links between specimen data and other relevant information, such as published literature, genomic variation, and environmental monitoring, would provide a more valuable and accessible resource for understanding and maintaining biodiversity.

1.6. Designate and encourage the development of a comprehensive, publicly accessible registry of United States–based biocollections, their holdings, and staff. Having such a resource will facilitate efforts to identify collections that are not yet part of the digitization community; to overcome the impediments that currently limit their involvement; to enlist participation by the broadest possible range and number of institutions; and to share information about the nationwide digitization effort, including standards, techniques, and funding opportunities. This activity should be done in coordination with existing efforts to create such a registry, such as Index Herbariorum and Scientific Collections International.

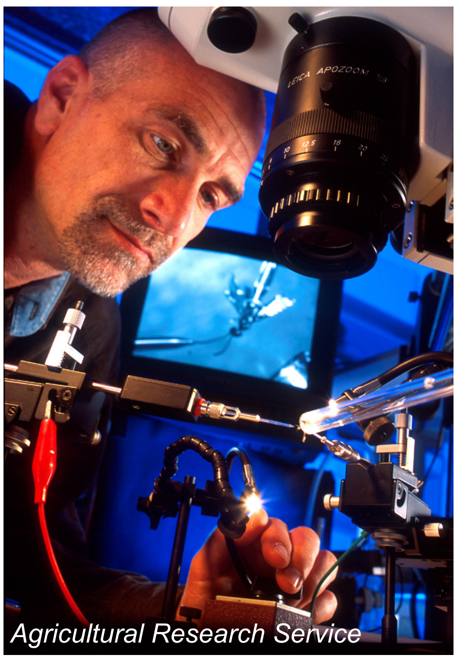
1.7. Create incentives for innovation in biocollections digitization. Institute a prize and other professional reward programs to provide monetary incentives and professional awards for developing ideas, tools, or other products that would make the digitization process faster, cheaper, and more accurate. The intent of incentives would be to attract individuals or teams with the professional software- or hardware-engineering expertise that NIBA needs to achieve its goals but that the collections community generally lacks.

1.8. Support existing efforts by host institutions and professional organizations (e.g., SPNHC, NSCA) to highlight the need for continued growth, improvement, and professional curation

of biocollections. NIBA embraces the rigor and power of a distinctive model for building databases and cyberinfrastructure, wherein data are fully validated by intact physical specimens. Specimen data can be verified, experiments can be replicated, and specimens can be further investigated in ways not possible when physical voucher specimens are unavailable. Therefore, the corresponding physical collections, which are the result of long-term efforts to build scientific infrastructure for investigations of the biosphere, must be properly maintained and must remain accessible for future discovery.

Goal 2: Advance engineering of the US biodiversity collections cyberinfrastructure. Implement adaptive technology strategies around core discipline standards to enable efficient digitization workflows, effective data management, permanent data archives, innovative and synthetic research, effective biodiversity policy, and ubiquitous educational engagement.

The Importance of Goal 2 for NIBA



Agricultural Research Service

Identification and inventory of the nation's biological diversity have always been a distributed and collaborative enterprise, but data networks have collapsed the spatial isolation of specimen repositories that are associated with centers of documentation, research, and education throughout the nation. Indeed, technology standards and especially data integration protocols have begun to reify hundreds of individual collections into a single data community by mobilizing and aggregating previously independent data stores. The resulting networked databases have led to the rediscovery of much of the ecological, historical, and taxonomic information associated with biological specimens. Going forward, cloud and high-performance computing, wireless communications, Web service architectures, and visualization on Web-based and mobile platforms promise additional transformational gains in multidisciplinary science integration and modeling (e.g., in the biological and Earth sciences) and in educational outreach.

There remain, however, significant technological gaps in the nation's cyberinfrastructure for biological diversity. National coordination is needed to harden key technology underpinnings that are vital to creating and sustaining an adaptive, integrative, and interoperable biodiversity data network. Filling those gaps with standard specimen data formats, supported network interfaces and protocols, and information semantics will yield a renaissance of research discovery and synthesis for biodiversity informatics. Finally, routine use of efficient workflows for data capture can insure the acquisition and dissemination of digital data from newly acquired specimens as they are incorporated into collections, thereby preventing new backlogs of nondigitized specimens from developing in the future.

The following implementation objectives will catalyze the required technological transformation.

The Implementation Objectives for Goal 2

2.1. Create a national database of all digitized specimen records from US institutions and agencies. Build an online database that provides access to all specimen records from the more than

1600 US biocollection institutions as these records are digitized. This national resource should be configured so that it can respond immediately to any biodiversity data query, whether from a scientist or a lay citizen. Through robust Web or programmatic interfaces, it should be easy to use by non-scientists and it should facilitate data integration and reuse by broader research, education, public health, and environmental agencies and by businesses. The following actions are required:

- Using existing national and international specimen data aggregation projects as models, specify the functional requirements of an aggregated US specimen data store.
- Consider existing implementation strategies that allow for significant technological innovation with storage, hosting, and cloud services to design a national database implementation strategy.
- Incrementally implement that strategy, providing technical support and protocols for ingesting specimen data records and associated images or other digital media and for the maintenance of these digital data by their source collections.
- Design protocols that are both robust and flexible and have the ability to accommodate taxon-specific and collection-type-specific differences in specimen distribution and preservation, data capture and curation, and so on.

2.2. Establish a research and development environment to deliver new specimen digitization workflow methods, tools, and techniques. Increasing the diversity and number of digitization mechanisms beyond those currently available will allow the biocollections community to take advantage of new hardware and software technology developments that will streamline and expedite data assimilation. The involvement of additional stakeholders will lead to the creation of technologies for crowdsourcing and for commercial open-source, open-design development. The following actions are required:

- Fill gaps in primary biodiversity data processing and workflows by creating software environments that take a beginning-to-end data processing perspective. Link and automate tasks, beginning with data acquisition in the field and ending in database assimilation, publication, and model integration.
- Derive a national consensus on specimen digitization goals and priorities and apply resources to develop specific technologies that will attain them. For example, a priority for entomology might be to design and deploy robotic (automated) 3D digital imaging instruments in sufficient numbers to digitize one million type specimens within 5 years and three million within 10 years.
- Develop technology support mechanisms that assist in digitization by institutions of all sizes, including the provision of easy-to-use and well-supported digitization software toolboxes and help-desk support.
- Develop professional infrastructure for the exchange and publication of digitization workflow methods, standards, and optimizations. Connect digitization tools and environments to online metadata archives for the automated capture and documentation of workflow methods, steps, costs, errors, scope, and so on. These metadata should be captured as part of online publications that document workflow design and optimization.

2.3 Complete development of required standards and protocols. These are needed to integrate primary specimen data with the interfaces, standards, and semantics of other environmental research communities and digital libraries, as well as with broader educational and commercial applications.

- Encourage informatics developers, including both nonprofit and commercial software providers, to publish open-source libraries and standards that link specimen data in collections with publications in online journals and open libraries of biodiversity publications. This includes the deployment of digital tools that pipe information from newly published species descriptions to online taxonomic databases, such as the Catalogue of Life, ZooBank, and

[Encyclopedia of Life](#) (EOL).[5] Particularly important is the choice and application of globally unique identifiers, mutual protocols, and common linking and resolving mechanisms. These and related decisions should be informed by consultation with collections managers and researchers.

- Engage the broader geosciences research community to unify standards for Earth science data, particularly as they apply to climatic, substrate, lithographic, geostratigraphic, and paleontological taxon data, and integrate these standards with those used for neontological specimens.

2.4. Promote a consensus for the adoption of standards. Make the standards developed for network protocols and computational interfaces straightforward and effective in order to allow seamless data transfer across computational environments of other research disciplines, government agencies, and educational and commercial organizations.

- Establish a working group or groups to identify constituent organizations for the current biodiversity architecture and to assess their roles, data sources, and tools, as well as the essential components missing from the architecture. The group or groups should also recommend a plan to add the missing components. The plan might include inviting additional participants, developing new tools, or other options for completing the architecture.
- Convene a working group to identify current problems and barriers to interoperability; develop standards for all aspects of digitization, including labeling, imaging, georeferencing, and required documentation; and develop system requirements for the biodiversity architecture.
- Develop methods for accreditation and promotion of software tools and network information or computation services that are technically robust, well supported, and actively maintained. Offer professional or prize incentives that endorse and promote software architecture and development projects that demonstrate a high level of usability, key digitization capabilities, data integration, and service interoperability among US biocollections.
- Establish a process for evaluating both existing and future digitization tools (from data capture to publication) that determines their ease of use and integration with other tools. The results will reveal the gaps in interoperability, which may then be addressed by focused tool development.
- Develop metric systems and adaptive analyses that maintain data integrity during and after the development of the knowledge system, including data quality and usability.

2.5. Anticipate the future of biodiversity specimen data integration.

- Convene a transformative and crosscutting scientific community panel that extends beyond disciplines and scientific genres to examine the process of integrating distinct information systems. The panel should define the process by which biodiversity collection information would be assimilated with the information bases from other science communities from genomics to Earth science.

2.6. Develop a strategy for long-term data archiving of specimen information, including 2D and 3D images, text information, and metadata about digitization processes.

- Develop, plan, and support long-term data archives that are perpetual and lossless through a consensus-based process (e.g., collective agreement or workshop) that focuses on expectations, strategies, and solutions for long-term data preservation.
- Explore existing models used by and possible collaborations with other ongoing data initiatives (e.g., digital libraries, [DataONE](#), iDigBio, NASA).

2.7. Support the development of a robust, Web-services-based architecture for handling taxonomic names applied to specimens as determinations and annotations.

- Complete discussions and planning of global names projects and bring the proposed data services, including knowledge representation environments and ontologies, into production computing environments.
- Complete standard protocols and network interfaces for digitization-client software and for taxon name servers, to develop client libraries and plugin modules for the open communication of names among database systems.

Goal 3: Enhance the training of existing collections staff and create the next generation of biodiversity information managers.

The Importance of Goal 3 for NIBA

Digitizing collections, as well as managing, editing, and sharing the data they provide, requires novel approaches to collection management and new skills for collections professionals. New career opportunities at the interface of biodiversity science and informatics are emerging, as are unprecedented opportunities for the general public to contribute to the scientific infrastructure. Because many US collections are international in scope, there are also significant opportunities to pursue international cooperative efforts in training in biodiversity science.


Florida Museum of Natural History

The following implementation objectives promote the development of a workforce that can lead and sustain the digitization effort.

The Implementation Objectives For Goal 3

3.1. Implement new training opportunities in biodiversity informatics.

- Develop new opportunities and expand existing training programs for collections professionals so that they can more fully engage in biodiversity informatics activities. These efforts should include exposure to informatics tools, their application both in biodiversity science and in informatics more broadly, and proper curation protocols for electronic data.
- Promote new undergraduate curricula and graduate programs in biodiversity informatics, particularly those that are cross-disciplinary (e.g., engineering, computer science, geography, library science). Develop opportunities for US students to gain international experience through biodiversity informatics training experiences using specimens and data that originated in other countries.
- Expand museum studies programs or biology degrees to include biodiversity informatics as applied to biocollections and exposure to topics such as informatics programming, visualization engineering, education and outreach visuals, natural language processing, and Web ontologies.

3.2. Establish career paths and professional retention incentives for data and specimen management and curation.

- Develop evaluation mechanisms that recognize and reward the products of successful careers in biodiversity informatics.

- Develop a standardized nomenclature and hierarchy for careers in biodiversity informatics that can inform position descriptions, hiring, and criteria for promotion.

3.3. Provide opportunities that encourage more people to become biodiversity software developers and that encourage the development of more biodiversity informatics software.

- Develop workshop training and software developer courses in biodiversity informatics.
- Develop schema that are accessible and establish service standards.
- Train experts and students in the various aspects of tool development and programming (e.g., MySQL).
- Collaborate with commercial firms to provide employment opportunities for biodiversity software developers who graduate from academic tracks.

Goal 4: Increase participation in and support for NIBA from a broad spectrum of stakeholders, both nationally and internationally.

The Importance of Goal 4 for NIBA


Scott Villa

The development of strategic partnerships with a large number of governmental and nongovernmental organizations is crucial to the success of NIBA. Such partnerships will help achieve a diversified funding base for the digitization effort, will ensure widespread use of the resulting data for both basic and applied research and education, and will create synergy with other initiatives of similar scope and goal.

The Implementation Objectives for Goal 4

4.1. Align NIBA's strategic plan with roadmapping exercises of other broad-based biodiversity initiatives in order to identify areas of overlap and synergy.

- Identify the position of collections data in the universe of other relevant data providers, share implementation plans, cross-report activities, and assemble a working group that represents related initiatives. Many current projects are both compatible with and complement NIBA in terms of its scope and potential impact. Examples include GBIF; BISON; Genomic Standards Consortium; Consortium for the Barcode of Life; ZooBank; Global Names Architecture; National Ecological Observatory Network; Earth Cube; Intergovernmental Platform on Biodiversity and Ecosystem Services; Convention on Biological Diversity; pro-iBiosphere; Scientific Collections International; and EOL. Each has objectives and activities that intersect at some level with NIBA's; coordination among projects is essential to minimize redundant effort and to maximize the use of available resources.
- Convene a workshop for representatives of the relevant professional societies and related groups to share digitization objectives, implementation strategies, and proposed activities. Such a workshop would reveal areas of mutual interest and overlap and would ultimately help distribute the effort required to attain common goals. It also would address the process

by which biocollections information should be assimilated with databases and other information sources from other science communities, from genomics to Earth science. Examples of relevant societies include SPNHC, NSCA, and AIBS, as well as taxon-specific societies.
- Establish convenient mechanisms (e.g., social media) whereby all members of initiatives with common objectives may readily interact with one another to share ideas, techniques, and so on.

4.2. Develop industry partnerships. Businesses that can benefit from biocollections data should be encouraged to support the digitization effort. Industry partners could contribute solutions to technological challenges of specimen digitization or provide a steady revenue stream for the digitization and maintenance of digitized data by using either a fee-for-service or a subscription model. Private companies concerned with health care, pharmaceuticals, agriculture, cosmetics, flavoring and fragrances, forestry, bioremediation, biological engineering, and environmental impact assessments are among those that could derive crucial information from specimen data.
- Enlist the assistance of an agency, organization, or company that can provide a matchmaking service between the collection digitization effort and likely business partners and that can create opportunities for direct contact with the technology transfer division of each business.
- Solicit industry partners to support a prize for technological innovation relating to collections digitization or data management.
- Develop models through which private companies would provide tangible support for collections whose data they use, such as commissioning data digitization, fees for use, or subscriptions. Through their philanthropic divisions, explore the interest of private companies in contributing funds for professional traineeships, student internships, and other mechanisms to support the endeavor.

4.3. Broaden the use of biocollections data for research across scientific disciplines. Unlocking collections data through digitization will make available billions of new data points, which can be incorporated into a wide range of scientific studies. The inherent value of such information is not well appreciated by many scientific research areas that will benefit from its use. Increasing the use of specimen data in a wide range of scientific studies not only improves such studies, but it also helps justify the expenditure of funds needed for collections digitization, maintenance, and growth.
- Enlist the assistance of the NSF or other funding agencies to incentivize the use and support of collections in research proposals.
- Publicize the availability of digital specimen data through outreach to scientific societies. A possible model is the data-mining and species-distribution modeling symposium sponsored by the Tri-Trophic TCN to be held at the University of California, Riverside, in 2014 to foster interaction between the systematics and ecological research communities and to explore the TCN's database as a platform for instruction and inquiry.
- Provide a help-desk function for ad hoc questions about collections and collections data use in collaboration with research and public librarians and through public outreach organizations such as EOL. Part of the help-desk activities could be the development of a data-quality metric system to convey the confidence limits of data in the knowledge base, as well as their fitness for use.
- Develop a system to track the use of specimen-derived data and credit the collections that contribute. Ideally, such a system would generate impact statistics that, when they are fed back to the contributing institutions, would not only serve to justify the effort of maintaining digital and physical collections but would also inform future digitization projects.

- Through partnership with the geosciences community, increase the rate and scale of the digitization of paleontological collections. Such collections are seriously underrepresented as an online resource in comparison with living organisms. Greater attention to fossils will enable significantly more research applications, ranging from the biotic response to environmental change to biogeography and to evolution and biodiversity loss. Opportunities to support digitization of paleontological collections may be attractive to funding agencies that do not support the biological sciences per se.

4.4. Expand NIBA to include a comprehensive effort to digitize federal collections. Because both federal and nonfederal collections data are required for many scientific studies, these sources must be integrated seamlessly with regard to physical, digital, and policy issues. This activity could be facilitated, for example, by establishing regional networks that comprise both federal and nonfederal collections within a given geographic region, or networks linked because of like collection types, or because of common research goals. Linking digitization efforts across adjacent institutions also provides attractive opportunities to share equipment and staff expertise, thereby cutting costs. The resulting efficiencies of scale would benefit all collections, and ensure that all collections are able to participate in international, national, regional, and local research initiatives. We strongly recommend focused, high-level discussions among federal entities, including the Biodiversity and Ecosystems Informatics Work Group, IWGSC, and BISON, to plan and implement the federal component of NIBA, which would promote the digitization of the US government's collections.

4.5. Recruit nonfederal (US and foreign) government agencies to join NIBA. Although BISON is intended to address the specific needs of US federal agencies for access to collections data, a variety of state and local agencies, as well as those of foreign governments, can similarly benefit from electronic access to such data. These include state parks and conservation departments; agencies that oversee zoning and land use; and government-sponsored programs such as CONABIO (National Commission for Knowledge and Use of Biodiversity, Mexico), Canadensys (Canada), and speciesLink (Brazil), which seek data from specimens collected in their countries that are stored in United States–based institutions. These data are particularly important for documenting biotic responses to climate change, understanding animal-borne human pathogens, and predicting the spread of invasive species.

- Develop a comprehensive contact list of relevant state and local agencies that are potential partners with US federal and foreign government entities.
- Encourage and enable collections institutions on a state or regional basis to reach out to these agencies for an exchange and, hopefully, an alignment of biodiversity data sources and needs.
- Initiate contact between NIBA and other national initiatives through joint participation in meetings and workshops.

4.6. Collaborate with digitization efforts pursued by other countries and share US biocollections data internationally.

- Strengthen collaborations with non-US biodiversity scientists and apply their place-based expertise in georeferencing and regional taxonomy for their countries to the digitization of US collections. This should be a reciprocal activity, because many specimens collected in what is now the United States, especially during early expeditions, are housed in non-US institutions.
- Collaborate actively with transnational initiatives to achieve an integrated, global knowledge base for biodiversity that is seamlessly interoperable.

Goal 5: Establish an enduring and sustainable knowledge base.

The Importance of Goal 5 for NIBA

Fully implementing NIBA will create a world-class resource for scientists, educators, the lay public, and policymakers alike. However, if the knowledge base encompassed by NIBA cannot survive beyond its initial development, the investment will have been squandered. The process for creating NIBA must also ensure that it can be sustained over a long time horizon. Its existence should not depend solely on the changing political priorities or budgetary swings of governmental or private funding agencies. Factors that will ensure that NIBA endures include the following:

- Fostering the widest possible participation. Many players should have a stake in NIBA's long-term success.
- Highlighting the importance of and need to digitize biological collections across the country.
- Configuring NIBA as an essential resource that attracts an ongoing revenue stream because its products are highly valued and widely used.

Especially in the currently tight budget environment, barriers to each of these factors must be evaluated and addressed to sustain NIBA over the long term. Such barriers include reduced government and private funding for collections, insufficient institutional support, the large number of disparate collections and the lack of an effective mechanism to identify and coordinate them, a diminishing pool of taxonomic experts for many groups of organisms, limited recognition of the critical need for a national biodiversity knowledge base, and intellectual property considerations that pertain to data sharing and use.

The implementation objectives for this goal address each of these potential barriers. Together, they constitute a plan to achieve a national biodiversity resource that can remain vibrant and useful over many years.

The Implementation Objectives for Goal 5

5.1. Identify and assess alternative economic models for sustaining NIBA.

- Convene a panel representing cross-cutting disciplines to develop and assess economic model for sustaining NIBA, including strategies for both revenue enhancement and expense reduction and including relevant information from analogous national-level resources in other countries. Any viable model should ensure ongoing support for relevant collections infrastructure and operations, including specimen acquisition, digitization, data and specimen curation, and data preservation. Intellectual property and other legal considerations that pertain to data sharing and use should be evaluated, especially regarding the possible use of data products as a source of revenue.

5.2. Institute changes in federal grant policies to support NIBA objectives.

- Work with federal agencies that fund specimen-based biological research to develop an agreement or policy that would require researchers to provide in each grant proposal a data management plan that includes the digitization of collections data as well as safe-guarding of physical specimens; to digitize such data in accordance with community-wide standards; and to share such data with NIBA. Existing data standards [6], OMB Circular A-130, and pending recommendations from the US Interagency Working Group on Digital Data should be evaluated and considered.

5.3. Secure enduring institutional support for collections digitization through access to staffing (collections) and technical (informatics) support.

- Demonstrate the value of physical and digital collections to the institutions that maintain them.
- Encourage collections institutions to establish long-term (perpetual), robust, and cost-effective plans for data archives. (See also item 2.6 above.)

5.4. Create a technology vision plan for the next-generation digitization of NIBA.

- Commission a forward-looking report that addresses the next generation of digitization, as well as related long-term maintenance, archiving, and stewardship issues. At a minimum, the report should address future efforts, funding, and the status of collections digitization.

Goal 6: Infuse specimen-based learning and exploration into formal and informal education.

The Importance of Goal 6 for NIBA

Kathy Frame

Convenient and immediate access to digitized information on planetary biodiversity provides unparalleled opportunities in education at a critical time. Although not contemplated in the original strategic plan for NIBA, implementation of the plan's objectives will require and benefit from early outreach to educators to highlight and facilitate the use of biodiversity data and informatics in classrooms and informal educational settings of all sorts. The recent American Association for the Advancement of Science report on science education in the United States [7], as well as a related National Academy of Sciences report [8], recommend shifting instructional methods toward hands-on, experiential, and problem-based lessons. National efforts such as the Partnership for Undergraduate Life Sciences Education will benefit from the new bioinformatics resources provided by NIBA, which will stimulate student-driven exploration of biodiversity. Currently, informatics is an emerging field that allows the exploration of large data sets, such as those generated by collections digitization.

NIBA therefore intersects in several places with the broad objective to create a more scientifically literate public. The nation's biocollection institutions represent a primary data source for research projects by high school and university students. Carefully crafted exercises can teach younger children about biodiversity and how to interpret information derived from primary sources. Participation in the digitization effort by student interns or citizen scientists can teach a wide range of skills, including the use of digital technology to explore the diversity of life, the relationships among organisms, and the interactions between organisms and their environments. Closer linkage of the NIBA effort to curricular goals, worker retraining, and general natural history literacy will benefit all.

The Implementation Objectives for Goal 6

6.1. Implement methods that allow K–20 educators to use digitized collections data.

- Develop scalable and transportable instructional curricula and analytic tools that are focused on place-based, experiential training for K–20 students in traditional and emerging fields (e.g., bioinformatics, climate change, evolutionary genomics, molecular ecology).
- Develop remote tools (e.g., mobile applications) that allow students and the general public to explore and interact directly with museum databases.
- Develop an integrated and dynamic network of K–20 educators who use specimen-based lessons in formal coursework.
- Conduct outreach to students from underrepresented groups (e.g., ethnic, cultural, physically challenged) with an emphasis on issues relevant to their communities.
- Award challenge grants to institutions to develop curricula that involve digitization. Funds might be contributed, for example, by the educational products industry (e.g., textbook publishers).
- Partner with colleges, universities, libraries, museums, and others to develop physical and virtual exhibits as well as other kinds of public programming and outreach that make use of collections data.

6.2. Engage citizen scientists to generate public enthusiasm for natural history.

- On the basis of emerging models in other areas of science [9], citizen-science projects have great potential to help realize both the scientific and the educational objectives of NIBA. Partnering with existing citizen-science programs (e.g., Citizen-Science.org, EOL, Zooniverse) to implement projects that introduce the public to collections-based biology and that foster participation in biodiversity informatics is one way to jumpstart the process.
- Involve amateur naturalists and taxonomists in taxonomic identification and in generating digitized content.
- Develop naturalist certification programs (e.g., master naturalists) that include significant exposure to digitized natural history collections and encourage valuable contributions.
- Provide incentives and public recognition for participation in digitization efforts (e.g., badges or certificates).
- Partner with vocational training programs and related civic initiatives that could further engage the general public in natural history through collections digitization.

Staging Capabilities, Products, and Deliverables

The campaign to digitize the nation's collections is not contingent solely on technology, organizational development, education, training, or professional engagement. Nor is it the sole responsibility of collections institutions, government entities or commercial interests. Instead it will require commitment from all of these stakeholders. With resources for a multidimensional strategic plan and prioritized parallel implementation activities, the biocollections community will deliver a national computing architecture for the widespread mobilization and engagement of specimen information. That transformation will leverage and reward the nearly 250-year investment that US collections have made in the documentation of the nation's biological diversity. With digitization activities advancing to a new level of commitment and investment, the community strains for leadership and consensus building to prioritize and guide the national effort to bring and sustain the US biological collections data online.

The following table summarizes the complementary and synergistic activities needed to mount and sustain this cyberinfrastructure campaign.

Table 1. Implementation Objectives. Numbers in parentheses indicate the corresponding key goals.

Goal 1: Establish an organizational and governance structure that will provide the national leadership and decision-making mechanism required to implement NIBA and to fully realize its strategic plan.			
Implementation objective	***Actions required***	***Time course***	***Status***
Found the NIBA organization. (1.1)	Contract with an existing organization, professional society, etc., to organize and lead NIBA for an initial period of 1–2 years.	Committee created within 6 months.	Not yet started.
Establish a governance, administrative, and management structure for NIBA. (1.2)	Secure funding to create the legal entity, convene governance and membership meetings, and support stakeholder engagement activities and leadership staff.	Complete within 18 months.	Not yet started.
Establish automated methods and metrics for measuring progress made toward NIBA's digitization goals. (1.3)	Develop schema, software protocols, and software libraries to describe and communicate specimen database growth statistics.	Ongoing.	Not yet started.
Document workflow challenges and identify solutions for collections with specialized or particularly challenging specimens. (1.4)	Canvass the collections community to describe difficult challenges, facilitate research and implementation of efficient solutions.	Ongoing.	Documentation, discussion, and research underway at iDigBio.

Implementation objective	*Actions required*	*Time course*	*Status*
Organize and facilitate both public and professional outreach to communicate NIBA's goals, programs, and accomplishments, as well as its broader societal benefits. (1.5)	Organize public outreach program, marketing materials, to point out value and relevance of NIBA activities.	Ongoing.	Not yet started.
Create a comprehensive, up-to-date, and publicly accessible registry of United States–based biocollections, their holdings, and curatorial staff. (1.6)	Identify all biocollection institutions, obtain contact, biographical information on staff, metadata on collections holdings.	Discovery and documentation of all biocollections to be completed within 2 years.	A structure for the comprehensive registry is under development, leveraging earlier attempts.
Create incentives for innovation in biocollection digitization. (1.7)	Establish prizes or other professional rewards to incentivize the creation of tools to digitize better, faster, and cheaper.	Begin immediately.	Not yet started.
Support efforts by host institutions and professional organizations to build, improve, and curate collections. (1.8)	Request additional funding for development and long-term care of specimens. Implement plans to improve the health of the biocollections enterprise (specimens, people, data).	Ongoing, but renew commitment and develop plan immediately.	Individual collections and disciplinary networks aware of needs and problems; some status reports generated on collections.
Goal 2: Advance engineering of the US biodiversity collections cyberinfrastructure. Implement adaptive technology strategies around core discipline standards to enable efficient digitization workflows, effective data management, permanent data archives, innovative and synthetic research, effective biodiversity policy, and ubiquitous educational engagement.			
Create a national database of all digitized specimen records from US collections. (2.1)	Document requirements, specify design, incrementally implement storage, computational, information retrieval, and Web integration capabilities.	Staged development goals, with yearly assessment of progress.	Prototype under development by iDigBio. Taxon discipline aggregations exist, resourced by grant funds.
Establish R&D environment to deliver new specimen digitization workflow methods, tools, and techniques. (2.2)	Support development of projects and proposals to identify and fill gaps.	Optimized, working solutions for all collection types within 3 years, ongoing incremental improvement.	Research underway with ADBC and TCN workflow research activities and working groups; software and hardware development needed.

Implementation objective	*Actions required*	*Time course*	*Status*
Develop infrastructure for the exchange and publication of digitization workflow methods. (2.2)	Define discrete workflow modules with individual performance metrics. Identify standards-based metadata schema and online repository for workflow process metadata.	Complete within 2 years, with standard schema, standards for description, evaluation, and comparison.	Related workflow documentation activities underway at iDigBio. Candidate workflow metadata standards exist.
Develop needed standards and protocols. (2.3)	Establish a working group to guide this activity.	Complete within 2 years.	General awareness of scope and options. Activities underway in iDigBio working groups.
Complete and maintain standards and software for mapping specimen data with interfaces and semantics of other research communities and organizations. (2.3)	Analyze existing opportunities for cross-disciplinary specimen data engagement, document existing standards and interfaces.	Complete data schema mapping tools and communication protocol libraries for integrating specimen data.	Activities underway in iDigBio working groups.
Develop methods for accreditation and promotion of robust software tools, network computation services, and data integrity. (2.4)	Establish a community software evaluation committee to assess maturity and interoperability of available software.	Online assessment of relevant software properties within 2 years.	Not yet started.
Invent and facilitate the future of biodiversity specimen integration. (2.5)	Organize existing examples of leading-edge integration, identify likely new areas leading to research insight.	Create a vision white paper and update annually.	Not yet started.
Develop an implementation plan for long-term data archiving of specimen information, including 2D and 3D images and text information. (2.6)	Survey and estimate the volume of storage needed, identify existing storage options, project options with technology changes.	Identify national solution within 3 years.	Not yet started.
Support the development of a robust, Web services-based architecture for handling taxonomic names applied to specimens as determinations and annotations. (2.7)	Bring data services associated with global names projects into production computing environments to provide open communication of taxonomic names among database systems.	Complete within 3 years.	Initial efforts are being pursued by NSF-funded Global Names Architecture and related activities.

Implementation objective	*Actions required*	*Time course*	*Status*
Goal 3: Enhance the training of existing collections staff and create the next generation of biodiversity information managers.			
Develop training programs in biodiversity informatics for museum professionals. (3.1)	Develop curriculum, determine proficiency level and requirements for attainment.	Recommend significant training curricula within 3 years.	Ongoing training workshops sponsored by several entities, including iDigBio.
Promote new undergraduate curricula and graduate programs in biodiversity informatics, expand museum studies programs to include software engineering and informatics activities. (3.1)	Identify topics and curriculum and funding resources, leverage existing programs.	Significant program curricula and course offerings within 3 years.	Some planning and nascent efforts underway.
Establish career paths, retention incentives for data or specimen management and curation. (3.2)	Develop evaluation mechanisms as well as standardized position nomenclature and promotion path.	Establish guidelines within 2 years.	Not yet started.
Define career paths and stimulate institutional employment of professional software developers. (3.3)	Workshop to define professional roles, institutional benefits, and sources of software engineering staff.	Report produced within 2 years with community outreach and promotion of engineering roles.	Not yet started.
Goal 4: Increase participation in and support for NIBA from a broad spectrum of stakeholders, both nationally and internationally.			
Achieve buy in and participation from a broad spectrum of stakeholders for NIBA's vision, activities and services. (4.1)	Convene a workshop with representatives of these groups to share objectives, implementation strategies to reveal areas of common interest.	Establish technical software and data collaborations within 2 years.	Many collaborative efforts underway with individual institutions and projects.
Develop industry partnerships. (4.2)	Outreach meetings with potential industrial partners.	Ongoing.	Not yet started.
Promote the use of digitized specimen data in research. (4.3)	Enlist assistance of funding agencies to incentivize collections support and use, publicize the availability of specimen data to scientific societies, provide help-desk functions for ad hoc questions about how to use collections data.	Ongoing, with yearly assessment.	Significant progress already made as a result of NSF guidelines regarding data management; workshops organized by iDigBio and one or more TCNs.

Implementation objective	*Actions required*	*Time course*	*Status*
Create metadata infrastructure to track specimen data usage for automatic attribution and credit reporting for source institutions. (4.3)	Develop an information architecture for logging collection data object usage for databases, Web services, Web pages.	Within 5 years, usage reporting infrastructure should be operational.	Not yet started.
Form a partnership with the geosciences community to increase the rate of digitization of underrepresented paleontological specimens. (4.3)	Identify the full range of paleontological collections and develop a comprehensive plan for digitization, addressing special challenges.	Complete national strategy plan within 3 years.	Activities underway in iDigBio working groups.
Expand NIBA to include digitization of federal collections. (4.4)	NIBA and NSF outreach to federal interagency committees and agencies on leveraging NIBA resources.	Ongoing, with yearly assessment.	Not yet started.
Develop nonfederal collaborations with international, regional, state, and local agencies with an interest in species occurrence data. (4.5)	Outreach activities that market available collection data resources.	Ongoing, with yearly assessment.	Activities have been initiated at iDigBio.
Initiate international collaboration to deliver US collection data to a global resource. (4.6)	Document the role of US national specimen data standards, then design for international interoperability of specimen data objects beyond Darwin Core.	Ongoing, establish benchmarks and yearly assessment metrics.	Individual collections and disciplinary networks currently share data with the GBIF cache; activities have been initiated at iDigBio.
Goal 5: Establish an enduring and sustainable knowledge base.			
Identify and assess alternative economic models for sustaining NIBA. (5.1)	Convene a cross-cutting panel to lay out economic models for sustaining NIBA and address other relevant issues, such as intellectual property concerns.	Complete within 2 years.	Sustainability planning initiated by iDigBio.
Institute changes in federal grant policies to support NIBA objectives. (5.2)	Work with federal agencies that fund biological research to require that researchers include specimen digitization in data management plans, digitize data in accordance with standards, and share data with NIBA.	Ongoing, with yearly assessment.	Not yet started.

Implementation objective	*Actions required*	*Time course*	*Status*
Secure institutional level support for digitization and access to staffing for collections and technical informatics support. (5.3)	Stakeholder organizations will work with members to articulate compelling arguments for increased and sustained institutional support for digital curation of collections data.	Ongoing.	Ongoing, through existing organizations and iDigBio outreach working group.
Create a technology vision plan for the next generation digitization future of NIBA. (5.4)	Commission a future-looking study that addresses innovations in digitization, as well as in related long-term data storage, databases, archiving, and data stewardship.	Updated annually.	Not yet started.
Goal 6: Infuse specimen-based learning and exploration into formal and informal education.			
Implement methods that allow K–20 educators to use specimen data as an integral part of curricula in environmental science. (6.1)	Identify and initiate partnership opportunities with education researchers and teachers to develop modular, accessible curriculum material.	Ongoing.	Initial workshops held on revising undergraduate curricula (*AIM-UP!*).
Engage citizen scientists in digitization projects. (6.2)	Partner with existing citizen-science platforms and projects.	Successful and widespread citizen-science initiatives within 3 years.	Several public participation digitization projects initiated; activities underway in iDigBio working groups.

[1] Skog J, McCourt RM, Corman J. 2009. *The NSF Scientific Collections Survey: A Brief Overview of Findings*. US National Science Foundation.

[2] McDade LA, Maddison DR, Guralnick R, Piwowar HA, Jameson ML, Helgen KM, Herendeen PS, Hill A, Vis ML. 2011. Biology needs a modern assessment system for professional productivity. *BioScience* 61: 619–625.

[3] [SPNHC] Society for the Preservation of Natural History Collections. 2011. *Prospects for a Scientific Software Innovation Institute in Biological Collections Digitization: Interim White Paper, September 2011*. SPNHC. (7 December 2012)

[4] President's Council of Advisors on Science and Technology. 2011. *Sustaining Environmental Capital: Protecting Society and the Economy*. Executive Office of the President of the United States.

[5] Miller J, et al. 2012. From taxonomic literature to cybertaxonomic content. *BMC Biology* 10 (art. 87).

[6] [NITRD] The Networking and Information Technology Research and Development program. 2009. *Harnessing the Power of Digital Data for Science and Society*. NITRD. (7 December 2012)

[7] Brewer CA, Smith D. 2011. *Vision and Change in Undergraduate Biology Education: A Call to Action*. American Association for the Advancement of Science.

[8] National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[9] Newman G, Wiggins A, Crall A, Graham E, Newman S, Crowston K. 2012. The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment* 10: 298–304.

Appendix A. NIBA Implementation Plan Workshop Attendees

The following individuals participated in the Workshop to Produce an Implementation Plan for a Network Integrated Biocollections Alliance. The workshop was held 17–18 September 2012 at the Hyatt Dulles Hotel in Herndon, Virginia.

Workshop Writing Committee

James Beach
Assistant Director for Informatics
Biodiversity Institute
University of Kansas

Joseph Cook
Professor, Director, and Curator of Mammals
Museum of Southwestern Biology
University of New Mexico

Linda S. Ford
Director, Collections Operations
Museum of Comparative Zoology
Harvard University

Robert Gropp
Director of Public Policy
American Institute of Biological Sciences

James Hanken (workshop co-organizer)
Director
Museum of Comparative Zoology
Harvard University

Kathy Joyce (workshop facilitator)
M. Kathleen Joyce and Associates

Lucinda McDade (workshop co-organizer)
Director of Research and Chair of the Botany Department
Rancho Santa Ana Botanic Garden

Barbara Thiers
Director
William and Lynda Steere Herbarium
New York Botanical Garden

Workshop Participants

Charles Bartlett
Assistant Professor
University of Delaware

Neil Cobb
Director
Merriam-Powell Center for Environmental Research
Northern Arizona University

Christopher Dietrich
Systematic Entomologist
Illinois Natural History Survey
University of Illinois at Urbana–Champaign

Jose Fortes
Professor of Electrical and Computer Engineering and Computer Science
University of Florida

Sara Graves
Director
Information Technology and Systems Center
University of Alabama in Huntsville

Corinna Gries
Center for Limnology
University of Wisconsin–Madison

Gerald “Stinger” Guala
Director of the Integrated Taxonomic Information System (ITIS)
Director of Biodiversity Information Serving Our Nation (BISON)
DOI Representative and Data lead, OSTP, NSTC Interagency Working Group on Scientific Collections
United States Geological Survey

Michael A. Mares
Director and Joseph Brandt Professor
Sam Noble Oklahoma Museum of Natural History
University of Oklahoma

Richard McCourt
Associate Curator of Botany
Academy of Natural Sciences of Drexel University

Amanda Neill
Director of the Herbarium
Botanical Research Institute of Texas

Christopher Norris
Senior Collections Manager
Peabody Museum of Natural History
Yale University

Larry Page
Curator of Fishes
Florida Museum of Natural History
University of Florida

Cynthia Parr
Director, Species Pages
Encyclopedia of Life
Smithsonian Institution

Greg Riccardi
Director
Institute for Digital Information and Scientific Communication
Florida State University

Nelson Rios
Manager of Collections and Bioinformatics
Tulane University

Katja Seltmann
Project Manager, Tri-Trophic Database Thematic Collections Network
American Museum of Natural History

Dena Smith
Curator, Invertebrate Paleontology
University of Colorado Museum
University of Colorado

Quentin Wheeler
Professor
School of Life Sciences
Arizona State University

National Science Foundation Observers

Melissa Cragin
American Association for the Advancement of Science, Science and Technology Policy Fellow
National Science Foundation

Daphne G. Fautin
Program Director, Division of Biological Infrastructure
National Science Foundation

Anne Maglia
Program Director, Division of Biological Infrastructure
National Science Foundation

Joann Roskoski
Deputy Assistant Director for Biological Sciences
National Science Foundation

Judith Skog
Expert, Advancing Digitization of Biological Collections
National Science Foundation

Appendix B. Solicitation for Comments on the Draft NIBA Implementation Plan

The NIBA Implementation Plan was developed through an iterative process. The Writing Committee prepared a draft report based on the discussions and recommendations of the participants in the Workshop. A draft report was then shared with Workshop participants and their comments were incorporated into a final draft report that was disseminated broadly for public comment. The following list summarizes the collections, biodiversity, and other stakeholder organizations and communication networks from which the Writing Committee sought comments.

List serves

- Ecolog-L (16,000 subscribers)
- Herbaria (698 subscribers)
- NHColl-L (1,322 subscribers)
- Taxacom (2,077 subscribers)

Organizations, Groups, and Other News Channels

- American Institute of Biological Sciences (160 member organizations)
- AIBS Public Policy Report (3,997 subscribers)
- AIBS_Policy Twitter (197 followers)
- AIBS Web site
- Editorial in the August issue of the journal *BioScience*
- iDigBio newsletter (435 people)
- Natural Science Collections Alliance Web site
- Natural Science Collections Alliance Washington Report (273 subscribers)
- Natural Science Collections Leadership Group on LinkedIn (253 members)
- Society for the Preservation of Natural History Collections members
- Federal agency and non-profit contacts who expressed interest in a NSC Alliance briefing on digitization (123 people)
- Chairs of biology departments at 125+ universities (300 people)

Suggested Citation:

American Institute of Biological Sciences. 2013. Implementation Plan for the Network Integrated Biocollections Alliance. Reston, VA, USA.