



BIODIVERSITY COLLECTIONS NETWORK

NIBA Status Report: Assessments and Recommendations from Biodiversity Collections Network Workshop

January 5-6, 2017

Washington, DC

Background

- I. The digitization effort, first outlined in a *Strategic Plan for a Network Integrated Biocollections Alliance (NIBA)* put forth a 10-year vision to
 1. Digitize data from all US biological collections, large and small, and integrate these in a Web-accessible interface using shared standards and formats;
 2. Develop new Web interfaces, visualization and analysis tools, data mining, and georeferencing processes and make all available for using and improving NIBA resources; and,
 3. 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.

- II. The *Implementation Plan for NIBA* (2013) identified six implementation goals:
 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.
 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.
 3. Enhance the training of existing collections staff and create the next generation of biodiversity information managers.
 4. Increase participation in and support for NIBA from a broad spectrum of stakeholders, both nationally and internationally.
 5. Establish an enduring and sustainable knowledge base.
 6. Infuse specimen-based learning and exploration into formal and informal education.

- III. Assessment of Progress and Recommendations
 1. The Biodiversity Collections Network, a National Science Foundation (NSF)–funded Research Coordination Network (RCN) project working to expand the

user community for digitized biocollections data and to help increase the sustainability of national digitization efforts, convened a workshop in January 2017 to gather community input on progress toward NIBA goals.

2. As part of the workshop, participants assessed progress toward these goals and offered recommendations for next steps.
3. This document summarizes the deliberations of the participants in this workshop.
4. Some goals below do not have detailed assessments or recommendations as the individuals participating in the workshop were not able to specifically address these goals. BCoN welcomes community input on these items, as well as on the other assessments and recommendations.

NIBA Implementation Goals—Summary

Implementation 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 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.

Implementation Goal 3:

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

Implementation Goal 4:

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

Implementation Goal 5:

Establish an enduring and sustainable knowledge base.

Implementation Goal 6:

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

Status Report on NIBA Goals and Workshop Comments and Recommendations

Implementation 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.

GOAL 1 STATUS REPORT

Accomplishments to Date:

Goal 1.1: Found the NIBA organization

- Following publication of NIBA Implementation Plan, representatives of the American Institute of Biological Sciences (AIBS), Society for the Preservation of Natural History Collections (SPNHC), Natural Science Collection Alliance (NSCA), and iDigBio met to consider next steps. Led by AIBS and with the active participation of the NSCA and SPNHC, the groups successfully submitted a Research Coordination Network proposal to NSF. A 5-year grant was funded in 2014 (NSF DBI-1441785).
- To date, the RCN has functioned as the NIBA organization, in loose collaboration with iDigBio—an NSF-funded National Resource addressing many NIBA goals (NSF DBI-1115210 and DBI-1547229).
- More than 20 organizations have endorsed the concept of NIBA via their formal endorsement of BCoN.

Goal 1.2: Establish a governance, administrative, and management structure for NIBA

- This has not yet been done.

Goal 1.3: Establish automated methods and metrics for measuring progress made toward NIBA's digitization goals

- iDigBio and NSF's Advancing Digitization of Biodiversity Collections (ADBC) program monitor digitization progress for those institutions funded by ADBC.
- Individual Thematic Collections Networks (TCNs) track progress toward digitization within their network.
- Beginning in 2017, *Index Herbariorum* will allow herbarium collections to self-report on their digitization progress.

Goal 1.4: Document workflow challenges and identify solutions for collections with specialized or particularly challenging specimens

- iDigBio has documented workflows for the digitization of various types of collections (<https://www.idigbio.org/content/digitization-workflows> and https://www.idigbio.org/wiki/index.php/Digitization_Resources).
- iDigBio has held multiple workshops to demonstrate digitization techniques and efficient workflows, and has held webinars to explore various tools for data capture and formatting (https://www.idigbio.org/wiki/index.php/IDigBio_Workshops).

Goal 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.

- BCoN organized a workshop to identify current communications and outreach resources and opportunities for increased community collaboration. A report from this workshop included specific actions for the community. These are now being developed and implemented via a Communications Working Group.
- A recommendation from the BCoN Communications Workshop was that biodiversity

collection professionals needed to become more skilled with engaging non-technical audiences. In response, AIBS has developed a Communications Boot Camp for Scientists. Biodiversity scientists and collections staff have and continue to participate in these programs, further increasing the capacity of the community to engage new audiences.

- Presentations about BCoN and its objectives have been given at several meetings of professional societies.
- WeDigBio holds citizen science events to spur global public engagement in digitization activities: www.wedigbio.org.
- Notes from Nature provides an attractive and engaging platform for transcribing data from natural history specimens, including blog posts that explain the importance of natural history collections: www.notesfromnature.org.

Goal 1.6: Create a comprehensive, up-to-date, and publicly accessible registry of United States–based biocollections, their holdings, and curatorial staff.

- GRBio/GRSCiColl created the structure for a worldwide registry of collections: www.grbio.org.
- *Index Herbariorum* has been funded to expand the information collected from herbaria to include information about digitization, and also to reach out to all the world’s herbaria for updated information: <http://sweetgum.nybg.org/science/ih/>.
- iDigBio curates a comprehensive list of all natural history collections in the United States of America -- <https://www.idigbio.org/portal/collections>.

Goal 1.7: Create incentives for innovation in biocollection digitization.

- ADBC program recommends that TCN proposals include innovations that could benefit digitization efforts for all institutions
- NSF and AIBS jointly administered the Beyond the Box prize challenge, a \$1M prize competition seeking to incentivize a group to create an automated method for digitizing insect drawers. No successful solution was identified.

Goal 1.8: Support efforts by host institution and professional organizations to build, improve, and curate collections.

- The Natural Science Collections Alliance (NSCA) has developed a new organizational strategic plan to help it focus its efforts and those of its institutional leaders on this task.
- AIBS, BCoN, SPNHC, iDigBio, California Academy of Sciences and NSCA have surveyed research collections to gather data about budgets, staffing, and priorities. A report is being written and should be submitted for publication by the end of 2017 or early 2018.
- AIBS, SPNHC, NSCA and others urged the NSF to reverse its decision to cut the Collections in Support of Biological Research (CSBR) program
- SPNHC, NSCA, professional societies and NSF have been quick to respond to collection emergencies, such as the transfer of specimens from University of Louisiana at Monroe to other institutions
- SPNHC is providing professional development through the emerging professional program that is engaging the next generation of biocollections staff professionals.
- SPNHC best practices wiki (<http://spnhc.biowikifarm.net/wiki>) is a depository for

information about curation and collections development.

GOAL 1 COMMENTS AND RECOMMENDATIONS FROM WORKSHOP

Restatement of Goal:

Building on existing organizational capacity, formalize a structure that provides the innovative leadership required to advance the updated NIBA vision, to address emerging biodiversity challenges.

Proposal for the Future:

Create four virtual or physical centers as described below, all overseen by a coordinating body, as outlined in the table below.

<p>Coordination Hub</p> <p>Does not yet exist. Primary role is coordination of the work among the four pillars. Representatives from all of the pillar institutions would need to be directly involved. Possible institutional homes could be within one of the pillar organizations or a separate, independent “center” within a host non-governmental organization</p>			
<p>Digitization Hub</p>	<p>Info Development Centers</p>	<p>Advocacy & Communications</p>	<p>Research Synthesis Center</p>
<p><u>Roles:</u> Implementing ADBC by serving as a data aggregator, developer of data standards and training, and leading outreach to promote use of the aggregated data.</p> <p><u>Education Elements:</u> Workforce training, undergraduate research opportunities, K–12 modules, citizen science engagement in</p>	<p><u>Roles:</u> Front-line implementers. Primary sources of data and institutional repositories for the original specimens. Source of expertise about the specimens and their suitability for various research efforts.</p> <p><u>Education Elements:</u> Direct engagement of citizen scientists in curation/digitization, deployment of specimen-based education to meet host institution needs (e.g., exhibits, school programs, student training).</p> <p><u>Current Orgs.:</u></p>	<p><u>Roles:</u> Provide a coordinated voice for the community, and advocate for its needs with funders and key stakeholders. Convener of the community via workshops to develop consensus strategies and proposals. This pillar would take the lead in coordinating the Communications Advisory Panel as proposed in Goal 4.</p> <p><u>Education Elements:</u> Development of marketing & communications materials and training for use by other pillar</p>	<p><u>Roles:</u> Advance cutting-edge research, coordinate inter-disciplinary research efforts, and publish recommendations for advancing research methods. Innovative roles could include documenting research successes and impact for inclusion in marketing and communications materials, and facilitating connections with the global research community.</p> <p><u>Education Elements:</u> Workforce training, research opportunities</p>

digitization efforts. <u>Current Org.:</u> iDigBio will play this role for the next four years and possibly continue to advance key aspects of this role beyond the end of its current grant.	The primary holders of specimen collections (museums, herbaria, universities, etc.) will play this role, and must be able to sustain their efforts. Data-development networks (e.g., TCNs) will contribute to this role, but will not necessarily be sustained over the long term.	institutions. <u>Current Orgs.:</u> AIBS and its member societies currently play this role, although only two years remain on the current RCN, and funds are insufficient to provide a sustained coordination effort, especially for marketing and communications.	for undergraduate and graduate students, development of K–12 curricula, and engagement of citizen scientists <u>Current Org.:</u> Does not yet exist. The NIBA community will develop a proposal for this center.
Goals 2, 3/6, and 7	Goals 2, 3/6 and 7	Goal 4	Goals 3 and 6

Immediate Action Items:

- AIBS will need to pursue funding for additional workshops to drive forward the coordinated communications needs. Longer term, a diversified funding portfolio of funding that includes private grants as well as support from the Information Development Centers will need to be developed.
- The BCoN team will need to form a proposal committee to follow up with NSF on possibilities for funding the creation of the center. BCoN should also look into the feasibility of inter-agency funding models (e.g., the International Cooperative Biodiversity Groups) that could be maintained over a longer time frame.

Implementation 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.

GOAL 2 STATUS REPORT

Goal 2.1: Create a national database of all digitized specimen records from US collections; document requirements; specify design; and incrementally implement storage, computational, information retrieval, and Web integration capabilities.

- iDigBio data portal currently contains >104 million records (includes data beyond that held at U.S. institutions): <https://www.idigbio.org/portal>.
- Regional or taxon-specific aggregations exist; most are duplicated in the iDigBio data

portal, and many are in GBIF as well. (For networks, see <http://usvhproject.org/#/networks>, though out of date. Also: [http://symbiota.org/docs/symbiota-introduction/active-symbiota-projects/.](http://symbiota.org/docs/symbiota-introduction/active-symbiota-projects/))

- BISON database is a unique, Web-based federal mapping resource for species occurrence data in the United States and its territories, with 375 million occurrence records, mostly insects: <https://bison.usgs.gov/#home>

Goal 2.2: Establish R&D environment to deliver new specimen digitization workflow methods, tools, and techniques; develop infrastructure for the exchange and publication of digitization workflow methods.

- NSF's Advances in Biological Informatics program (ABI) is supporting the development of Kurator, a project to develop a provenance-enabled Workflow Platform and Toolkit to Curate Biodiversity Data: [http://wiki.datakurator.org/wiki/.](http://wiki.datakurator.org/wiki/)
- iDigBio has documented workflows for the digitization of various types of collections -- <https://www.idigbio.org/content/digitization-workflows> and https://www.idigbio.org/wiki/index.php/Digitization_Resources.

Goal 2.2.1: Support development of projects and proposals to identify and fill gaps.

- No information available.

Goal 2.2.2: Define discrete workflow modules with individual performance metrics. Identify standards-based metadata schema and online repository for workflow process metadata.

- No information available.

Goal 2.3: Develop needed standards and protocols; establish a working group to guide this activity; complete and maintain standards and software for mapping specimen data with interfaces and semantics of other research communities and organizations.

- Digitization best practices: https://www.idigbio.org/wiki/index.php/Recommendations_for_the_Acquisition,_Processing,_and_Archiving_of_Digital_Media.
- Standards developed by TDWG community, such as Darwin Core and Audubon Core; improvement of collections management system platforms such as Symbiota to meet needs of collaborative digitization efforts: [http://www.tdwg.org/standards/.](http://www.tdwg.org/standards/)
- There have been many workshops funded through iDigBio to share best practices and efficient workflows for digitization: https://www.idigbio.org/wiki/index.php/Digitization_Resources#iDigBio_Workshops_2C_Reports.2C_and_Wikis.
- The practice of assigning identifiers is becoming more standard in specimen databases and is accommodated by the iDigBio database and other aggregators.
- Genbank/NCBI is providing voucher linkage to specimen records while collections databases are beginning to record Genbank sequences associated with specimens.
- The NSF-funded Wholetale initiative (www.wholetale.org) is looking at mechanisms to merge science and cyberinfrastructure pathways while the Pensoft data integration tool (<http://arpha.pensoft.net/>) is enabling the direct publication of specimen

examined data from iDigBio, GBIF, and DataOne. BiSciCol (<http://biscicol.blogspot.com/>) is also tackling the problem of interoperability of disparate data sources.

- Symbiota and Specify open-source collections management platforms now have over 350 and 500 collections, respectively, using and contributing to their software. They both present online and recorded workshops, training modules, and webinars on effective use and maintenance and have large online resources of training documentation. A large number of the TCN's are using these software packages to manage the data being generated and served to the community.

Goal 2.4: Develop methods for accreditation and promotion of robust software tools, network computation services, and data integrity; Establish a community software evaluation committee to assess maturity and interoperability of available software.

- No information available.

Goal 2.5: Invent and facilitate the future of biodiversity specimen integration; organize existing examples of leading-edge integration; identify likely new areas leading to research insight.

- Both the EarthCube RCN and the Paleobiology Database Project (<https://paleobiodb.org>) have made progress in this area along with various paleo and Earth science-specific workshops and working groups through iDigBio. There is also a paleo extension to Darwin Core that allows for better mapping of paleo-specific data. Some of the existing models include Dryad, DataOne and iDigBio. iDigBio has also addressed this with their Data Management Working Group¹
- BiotaPhy is a project to integrate research pipelines to link data, CI, and analytical tools established by recent NSF investments in biodiversity science—the Open Tree of Life, iDigBio, Lifemapper, and tools for comparative methods in Arbor. For example, Arbor (arborworkflows.com/; Harmon et al. 2013), hopes to provide streamlined, powerful workflows for evolutionary analyses. Although workflows/linkages are in development, the underlying biodiversity programs developed by diverse authors as R packages are already available for use: http://wiki.biotaphy.org/index.php?title=Main_Page.

Goal 2.6: Develop an implementation plan for long-term data archiving of specimen information, including 2D and 3D images and text information; survey and estimate the volume of storage needed; identify existing storage options, project options with technology changes.

- The Cyverse data store has been used as an archive for several TCN projects: www.cyverse.org
- Morphosource is a project-based data archive that allows researchers to store and organize, share, and distribute their own 3D data, and download and print 3D fossils: www.morphosource.org.
- Zenodo is a research data repository. Launched in 2013, it allows researchers in any subject area to upload files up to 50 GB: <https://zenodo.org>.

¹ https://www.idigbio.org/wiki/index.php/Data_Management_Interest_Group

- Data Observation Network for Earth (DataONE) is a community driven project providing access to data across multiple member repositories, supporting enhanced search and discovery of Earth and environmental data. DataONE promotes best practices in data management through responsive educational resources and materials: www.dataone.org.

Goal 2.7: Support the development of a robust, Web services based architecture for handling taxonomic names applied to specimens as determinations and annotations; Bring data services associated with global names projects into production computing environments to provide open communication of taxonomic names among database systems.

- The Global Names Architecture project has produced a Global Names Resolver: <http://resolver.globalnames.org/>.
- Cyverse has funded the Taxonomic Names Resolution Service for plants: <http://tnrs.iplantcollaborative.org>.
- Kurator: <http://wiki.datakurator.net/web/Kurator>².
- Ontobrowser: <https://github.com/Novartis/ontobrowser>.
- Animal Diversity Web: <http://animaldiversity.org/site/about/technology/index.html/>.
- Biodiversity Collections Ontology (BCO): <https://github.com/BiodiversityOntologies/bco>.

GOAL 2 COMMENTS AND RECOMMENDATIONS FROM WORKSHOP

Proposal for the Future:

Continue the following existing activities:

- Continue digitization of existing collections and field notes, journals, slides, vocalizations, and other forms of data.
- Continue to encourage collections to share data through aggregators.
- Adoption of standards and best practices.
- Lead collaborative efforts with other national, including federal, and international sources of data to enable data integration.
- Encourage the enrichment of data by collections (e.g., through the use of field notes) and by the researcher community through their routine utilization of specimens.
- Coordinate with the research community to expand and enrich the knowledge base by reporting use of collections, returning collections products derived from specimen based research as well as expanding on their collecting activities to obtain more material.

Implement the Following:

- Develop annotation systems and other methods for data enrichment to improve quality and utility of data.
- Advance creation of tools through aggregators for the investigation of comparative collection statistics only possible at the aggregator level - measures of uniqueness, data cleanup, visualization tools, and collection use statistics.

² <http://www.sciencedirect.com/science/article/pii/S1877050912002980>

- Design models for facilitating digitization, aggregation and linking of data obtained from multidisciplinary collecting endeavors.
- Investigate models for community long-term database file archiving.

Immediate Action Items:

- BCoN should convene a workshop to bring together all players in the data pipeline (collection databases, aggregators, publishers, repackagers, archives) to discuss existing hardware and software infrastructure, identify gaps, and implement strategies for closing those gaps.

Implementation Goal 3:

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

GOAL 3 STATUS REPORT

Goal 3.1: Develop training programs in biodiversity informatics for museum professionals; develop curriculum, determine proficiency level and requirements for attainment; promote new undergraduate curricula and graduate programs in biodiversity informatics; expand museum studies programs to include software engineering and informatics activities; identify topics and curriculum and funding resources, leverage existing programs.

- A course has been developed at the University of Florida that explores the use of collections data for research. Online enrollment in the course was enabled to increase access to the course.
- TCN projects have provided hands-on training for a significant number of undergraduate students.
- iDigBio has developed and continues to offer a series of workshops and training opportunities that focus on digitization of particular kinds of biological specimens, with emphasis on those that present special challenges.
- SPNHC Emerging Professionals Committee has developed a list of courses on various aspects of collections management that are routinely shared with students and early career professionals.
- Biodiversity Literacy in Undergraduate Education (BLUE) is an initiative to build partnerships among biodiversity and education researchers to identify strategies, centralize resources, and develop educational materials:
<http://www.biodiversityliteracy.com/blue-about>.

Goal 3.2: Establish career paths, retention incentives for data or specimen management and curation; develop evaluation mechanisms as well as standardized position nomenclature and promotion path.

- BCoN co-sponsored and participated in an AIBS-organized meeting addressing biological informatics education and workforce issues during which these issues were explored. A report (https://www.aibs.org/events/council_meeting_2015.html) from that meeting was shared with professional societies, data management and analysis organizations, and education organizations.

Goal 3.3. Define career paths and stimulate institutional employment of professional software developers; establish workshop to define professional roles, institutional benefits, and sources of biodiversity informatics staff.

- Monfils, A.K. 2015. Biological informatics training needs for systematics and digitization. Addressing Biological Informatics Workforce Needs: American Institute of Biological Sciences. 2015 AIBS Council Meeting, Washington DC.
- Crandall, K., R. Gropp, and A.K. Monfils. 2016. Addressing Biological Informatics Workforce Needs. American Institute of Biological Sciences Webinar Series.

GOAL 3 COMMENTS AND RECOMMENDATIONS FROM WORKSHOP

Restatement of Goal:

It was determined that this goal does not require restatement.

Proposal for the Future:

- Share job descriptions that are relevant to larger collections where specimen curatorial and data curatorial tracks may be differentiated, as well as other versions relevant to smaller collections where a single curator is likely to be responsible for all aspects of the endeavor.
- Share curricula and course content for museum studies courses and programs with the community at large. Ideally, some of these (or some sessions of some of these) would be captured as webinars and made available digitally.
- Distribute widely notice of educational opportunities that are available through iDigBio, the TCNs, and any other participating entity that is willing to share or make their opportunities available.
- Create a mentorship program wherein experienced curators would work with novice curators as apprentices to provide guidance to them to help them accomplish their curatorial responsibilities. This might involve travel by the apprentice to the mentor's home institution or could be via webinar or other digitally enabled vehicle(s).
- Develop a curator training certificate program.
- Develop best practices for museum studies curricula and mentorship practices.
- A Biodiversity Hub / Synthesis Center / Knowledge Center could serve as a clearinghouse for the items described above.

Immediate Action Items:

No immediate actions were identified.

Implementation Goal 4:

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

GOAL 4 STATUS REPORT

Goal 4.1: Achieve buy in and participation from a broad spectrum of stakeholders for NIBA's vision, activities, and services; convene a workshop with representatives of these groups to share objectives and implementation strategies to reveal areas of common interest.

- See BCoN (www.bcon.org) and iDigBio (www.idigbio.org) for lists of workshops, webinars, symposia, and other programs. There have been many per year since 2011.

Goal 4.2: Develop industry partnerships and outreach meetings with potential industrial partners.

- Picturae (<https://picturae.com>) has developed a business of high-throughput specimen digitization, especially for herbarium specimens.
- CISCO funded a 2-day "Pit Stop" meeting to bring collections professionals together with small to mid-sized technology companies at the Natural History Museum in London to explore ways to make specimen digitization more efficient—iDigBio participated in this meeting.

Goal 4.3: Promote the use of digitized specimen data in research; 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.

- BCoN established a Research Working Group, which participated in this workshop.
- NSF established a postdoctoral fellowship to support research using biological collections.
- This is a major current focus of iDigBio, with activities such as booths and professional society meetings, symposia at professional meetings, Research Spotlight (<https://www.idigbio.org/tags/research-spotlight>), and social media posts.
- The iDigBio data portal uses technology to make data accessible, such as specimen identifiers and APIs, and to answer questions from the community about how to use the data.
- Paleoniches TCN and iDigBio created a mapping plugin that can generate maps of fossil specimen localities (<https://github.com/iDigBio/wp-leaflet-idigbio-geojson-data>).
- TDWG Biodiversity Data Quality Interest Group is working toward developing standardized terminology and tools for assessing the research fitness-for-use for biodiversity data: <http://www.tdwg.org/activities/biodiversity-data-quality/interest-group-charter/>.
- A collaboration between iDigBio, Encyclopedia of Life (EOL), and Jorrit Poelen (freelance software engineer, GloBI) is resulting in two tools using iDigBio primary biodiversity data and other data sources: Effechecka is a taxonomic checklist generator which uses spatial and trait data, and FreshData enables researchers to receive notification as to when new biodiversity data, based on a particular query string, has become available from various data sources, such as iDigBio.
- iDigBio is working with the San Diego Super Computer facility to develop a SuAVE extension for curating and exploring large collections datasets, support domain standards, interface with taxonomies, and make them available through an analysis

platform. SuAVE uses technical approaches from image analytics, faceted search, and online map navigation as an online tool for exploratory data analysis.

- Pensoft's ARPHA publishing Platform has been developed to streamline the import of biodiversity occurrence records from iDigBio and other databases into the Biodiversity Data Journal: <http://arphahub.com/>.
- ePANDDA (Enhancing Paleontological and Neontological Data Discovery) is a new API being developed for geoscience database connectivity -- <https://epandda.org/>
- An NSF-funded project called "Connecting resources to enable large-scale biodiversity analyses"—which is known as BiotaPhy—is developing a cyberinfrastructure framework linking diverse data (phylogeny, morphology, ecology, fossils, geography, and climate) across several data-driven community research tools: http://wiki.biotaphy.org/index.php?title=Main_Page.
- PhyloJIVE (Phylogeny Javascript Information Visualiser and Explorer) is to be upgraded to Phylolink, which integrates taxon information across phylogeny, biogeography, geology, and paleobiology: <http://phylolink.ala.org.au/>. A joint project of iDigBio and Atlas of Living Australia.

Goal 4.4: Create metadata infrastructure to track specimen data usage for automatic attribution and credit reporting for source institutions; develop an information architecture for logging collection data object usage for databases, Web services, and Web pages.

- GBIF has developed tools for dataset metrics: <https://devpost.com/software/gbif-dataset-metrics-xfvzns>

Goal 4.5: Expand NIBA to include digitization of federal collections; NIBA and NSF outreach to federal interagency committees and agencies on leveraging NIBA resources.

- Nothing was reported.

Goal 4.6: Develop nonfederal collaborations with international, regional, state, and local agencies with an interest in species occurrence data; outreach activities that market available collection data resources.

- PhyloJive (Phylogeny Javascript Information Visualiser and Explorer), to be upgraded to Phylolink, integrates taxon information across phylogeny, biogeography, geology and paleobiology (<http://phylolink.ala.org.au/>). A joint project of iDigBio and Atlas of Living Australia.
- Workshop held in Frankfurt, Germany, in March 2017 explored the international collaboration between biodiversity information projects and infrastructure, and produced a broad commitment from all participants to try to develop better mechanisms to make this possible.
- The Global Biodiversity Informatics Outlook proposes a framework that will help harness the immense power of information technology and an open data culture, to gather unprecedented evidence about biodiversity and to inform better decisions: <http://www.gbif.org/resource/80859>.

- The GBIF Task Force, Accelerating the Discovery of Biocollections Data. This international task force sought to canvass the international collections community about the status and priorities for digitization, to document best practices from ongoing content mobilization, and to provide guidance on establishing priorities for digitizing biocollections: <http://www.gbif.org/resource/83022>.

Goal 4.7: Initiate international collaboration to deliver US collection data to a global resource; document the role of US national specimen data standards, then design for international interoperability of specimen data objects beyond Darwin Core.

GOAL 4 COMMENTS AND RECOMMENDATIONS FROM WORKSHOP

Restatement of Goal:

Expand the inclusiveness of holistic sampling to reflect life on Earth and increase the discoverability, visibility, and use of the resulting biodiversity collections with a broad spectrum of stakeholders, both nationally and globally.

Proposed activities for the Future:

- Integrate NEON, federal collections, and non-federal governmental collections.
- Create a resource portal.
- Develop a strategy for improving communication and integration between all aspects of collections, including scientists, collections curatorial staff, data managers, and all other involved in the collections process.
- Create fora for developing new partnerships across a range of actors, including public organizations, industry, scientific communities, and citizen scientists.
- Provide better outreach for understanding and articulating the value of biodiversity collections.
- Advocate for holistic biodiversity collections that support improved understanding of life on earth.

Implementation Goal 5:

Establish an enduring and sustainable knowledge base.

GOAL 5 STATUS REPORT

Goal 5.1: Identify and assess alternative economic models for sustaining NIBA; convene a cross-cutting panel to lay out economic models for sustaining NIBA and address other relevant issues, such as intellectual property concerns.

- Director's Summit sponsored by iDigBio at SPNHC 2017 in Denver brought together directors of collections institutions to discuss the value of digitization to institutions and strategies to sustain this effort.

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

- A Data Management Plan is now required as part of NSF proposals.

Goal 5.3: Secure institutional-level support for digitization and access to staffing for collections and technical informatics support; articulate compelling arguments for increased and sustained institutional support for digital curation of collections data; commission a future-looking study that addresses innovations in digitization, as well as in related long term data storage, databases, archiving, and data stewardship.

- BCoN convened a Communications Workshop, which has helped engage biodiversity collections professionals in communications training and to help them better understand how to engage novel communities.
- Director's Summit at iDigBio SPNHC 2017 in Denver brought together directors of collections institutions to discuss the value of digitization to institutions and strategies to sustain this effort.

Goal 5.4: Create a technology vision plan for the next generation digitization future of NIBA; commission a future-looking study that addresses innovations in digitization, as well as in related long-term data storage, databases, archiving, and data stewardship.

- The iDigBio External Advisory Board is providing iDigBio and ADBC with advice about long-term data storage, databases, archiving, and data stewardship.

GOAL 5 COMMENTS AND RECOMMENDATIONS FROM WORKSHOP

Restatement of Goal:

- Emphasize collaborative resource development.
- Make it clear that this goal is about developing resources (especially financial).
- Set the stage for developing strategies that develop resource for specific needs identified in the other goals (as opposed to generic strategies about developing funding from various sectors like corporations).

Proposal for the Future:

- Establish a Research Synthesis Center or other National Resource to address this goal.
- Develop a coordinated marketing and communications strategy.
- Secure an expanded Collections in Support of Biological Research (CSBR) grant program within NSF.

Immediate Action Items:

- Secure funding for a Research Synthesis Center.
- Secure funding for workshops to advance a coordinate marketing and communications strategy. Include participation of outside experts, and seek support for market research to help develop effective messages and branding.
- Request that NSF expand and increase investments in the CSBR program within NSF.

Implementation Goal 6:

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

GOAL 6 STATUS REPORT

Goal 6.1: Implement methods that allow K–20 educators to use specimen data as an integral part of curricula in science; identify and initiate partnership opportunities with education researchers and teachers to develop modular, accessible curriculum material.

- CollectionsWeb Research Coordination Network (2007–2013) identified grand challenges associated with integrating museum specimens into K-12 and undergraduate education and produced three manuscripts related to natural history collections in undergraduate education:
 - Powers, K. E., L. A. Prather, J. A. Cook, J. Woolley, M. Blackwell, H. L. Bart, Jr., A. K. Monfils, P. Sierwald. 2013. CollectionsWeb Workshop Report: New Approaches to Specimen-based Education. Unpublished report submitted to the US National Science Foundation.
 - Monfils, A.K., K.E. Powers, C.J. Marshall, C.T. Martine, J.F. Smith and L. Alan Prather. 2016. Natural History Collections: Teaching about Biodiversity across Time, Space, and Digital Platforms. *Southeastern Naturalist*. In press.
 - Powers, K.E., L.A. Prather, J.A. Cook, J. Woolley, H.I. Bart, Jr., A.K. Monfils, and P. Sierwald. 2014. Revolutionizing the use of natural history collections in education. *The Science Education Review* 13 (2): 24–33.

- Advancing the Integration of Natural History Collections: AIM-UP! Research Coordination Network - Undergraduate Biology Education (2010–2016) was effective at developing specimen based educational modules, advocating for using specimens and specimen based data in the undergraduate classroom, and growing a network of engaged faculty and collections professionals interested in integrating museums into undergraduate programs. A program evaluation was conducted and submitted to NSF, this includes an assessment of ongoing needs and future directions. Several publications resulted from the network.
 - Monfils, A.K. 2016. Advancing Integration of Museums into Undergraduate Education: AIM-UP! Program Evaluation. Unpublished report submitted to the US National Science Foundation.
 - E.A. Lacey, T.T. Hammond, R.E. Walsh, K. Bell, S.V. Edwards , E.R. Ellwood, R. Guralnick , S.M. Ickert-Bond, A.R. Mast, J. McCormack, A.K. Monfils, P.S. Soltis, D.E. Soltis, and J.A. Cook. 2017. Climate Change, Collections and the Classroom: Using Big Data to Tackle Big Problems. *Evolution: Education and Outreach*. In press.
 - Cook, J.A., E.A. Lacey, S.M. Ickert-Bond , E.P. Hoberg, K.E. Galbreath, K.C. Bell, S.E. Greiman, B.S. McLean, S.V. Edwards. 2016. From Museum Cases to the Classroom: Emerging Opportunities for Specimen-Based Education, Pp in Pavlinov, I., ed. “Aspects of Biodiversity” Volume 54 of the Archives of Zoological Museum of Moscow State University. In press.
 - Cook, J.A., S.V. Edwards, E. Lacey, R.P. Guralnick, P.S. Soltis, D.E. Soltis, C. Welch, K.C. Bell, K.E. Galbreath, C. Himes, J.M. Allen, T.A. Heath, A.C. Carnaval, K.L. Cooper, M. Liu, J. Hanken, and S. Ickert-

- Bond. 2014. Natural History Collections as Emerging Resources for Innovative Education. *BioScience* 64: 725–734.
- Ho, S-H. and J.A. Cook. 2013. Co-Evolving Pedagogies. *ARID: A Journal of Desert Art, Design and Ecology*. Vol. 2, published online
 - Dunnum, J. L., and J.A. Cook. 2012. Gerrit Smith Miller: His influence on the enduring legacy of natural history collections. *Mammalia* 76: 365-373.
- iDigBio Education and Outreach Working Group (2012–present) has a mission to coordinate education and outreach (E&O) activities among iDigBio, TCNs and other partnering projects, share E&O related resources, tools, and training with the ADBC community and beyond, and plan, develop and implement E&O activities of interest to the HUB and TCNs.
 - iDigBio Undergraduate Education Workshop: Building and Disseminating Resources for Collections-Based Undergraduate Education--
https://www.idigbio.org/wiki/index.php/Resources_for_Collections-Based_Undergraduate_Education
 - iDigBio K–12 Education Workshop: Incorporating K–12 Outreach into Digitized Collections Programs--
https://www.idigbio.org/wiki/index.php/Incorporating_K-12_Outreach_Into_Digitized_Collections_Programs_Workshop
 - 3D Digitization of Fossils for Educators & Citizen Scientists --
https://www.idigbio.org/wiki/index.php/3D_Digitization_of_Fossils_for_Educators_%26_Citizen_Scientists:A_collaborative_workshop_among_idigBio,_the_FOSSIL_Project,_and_K12_Science_Educators
 - Biodiversity Literacy in Undergraduate Biology: BLUE (2016 –present) formed to build from the AIM-UP! momentum, synergize efforts with the QUBES Hub, incorporate WeDigBio educational initiatives, and work with educational initiatives at iDigBio. This collaborative has submitted a successful workshop proposal to iDigBio for an education workshop in spring of 2017, outlined a workflow for developing educational modules using natural history collection data, produced an educational module for an introductory core curriculum bringing in partners from Esri, Kurator, and Ecological Society of America (ESA), and run workshops and developed faculty mentoring networks using museum data at both the BioQUEST, Science Case Net and QUBES-sponsored 2016 National Academies Special Topics Summer Institute on Quantitative Biology "*Lowering the Activation Energy: Making Quantitative Biology More Accessible*" summer institute 2017 and the *Resources for Ecology Education Fair & Share* (REEFS) workshop at ESA 2016. The goals of this group include bringing new partners together to advance undergraduate educational initiatives, combine resources, avoid duplicate efforts, and disseminate educational materials. This includes engaging new and early adopters of educational materials to support effective educational modules through implementation.
 - T. Barbaro, D. Bloom, J. Cook, S. Donovan, E. Ellwood, D. Linton, A.K. Monfils, M. Phillips, and J. Whorley 2016. Incorporating Digitized Natural History Collections Data into the Classroom. 2016 National Academies Special Topics Summer Institute on Quantitative Biology

- “Lowering the Activation Energy: Making Quantitative Biology More Accessible.” Raleigh, NC. (Resource Session and Poster Presentation).
- Monfils, A.K., E. Ellwood, D. Linton and M. Phillips. 2016. Evolution (and Co-Evolution) of Angiosperms: Collections in the Classroom. REEFS: Resources for Ecology Education, Fair and Share. 2016 Ecological Society of America Annual Meeting, Fort Lauderdale, FL
 - Biological Collections Network: BCoN Research Coordination Network (2014–present) has supported educational initiatives (both workforce training and undergraduate) that address the needs of the collection community. Support was provided for the BLUE collaborative to present at the 2016 National Academies Special Topics Summer Institute. Efforts include co-sponsoring the 2015 AIBS Council Meeting, in Washington DC, on biological informatics workforce training needs for systematics and digitization. This meeting resulted in a report and webinar.
 - Monfils, A.K. and S. Donovan. 2016. BCoN Workshop: Integrating Resources and Growing the Community: Data Resources and Data Literacy. Scheduled for November 17–20, 2016.
 - Barbaro, T., D. Bloom, J. Cook, S. Donovan, E. Ellwood, D. Linton, A.K. Monfils, M. Phillips, and J. Whorley 2016. Incorporating Digitized Natural History Collections Data into the Classroom. 2016 National Academies Special Topics Summer Institute on Quantitative Biology “Lowering the Activation Energy: Making Quantitative Biology More Accessible.” Raleigh, NC. (Resource Session and Poster Presentation)
 - American Institute of Biological Sciences. 2016. Addressing Biological Informatics Workforce Needs: A Report from the 2015 AIBS Council of Member Societies and Organizations Meeting. Reston, Virginia.
 - Monfils, A.K. 2015. Biological informatics training needs for systematics and digitization. Addressing Biological Informatics Workforce Needs: American Institute of Biological Sciences. 2015 AIBS Council Meeting, Washington, DC.
 - Crandall, K., R. Gropp, and A.K. Monfils. 2016. Addressing Biological Informatics Workforce Needs. American Institute of Biological Sciences Webinar Series.
- The BCoN RCN supported a working group meeting, titled "Integrating Resources and Growing the Community: Data Resources and Data Literacy." Anticipated workshop outcomes include a manuscript submission to *BioScience* on data literacy and workforce needs and a specific plan and timeline for creating a network of educators and data providers working together to further data literacy in the foundational undergraduate community.

Goal 6.2: Engage citizen scientists in digitization projects; Partner with existing citizen science platforms and projects.

- Notes from Nature (www.notesfromnature.org) —a crowdsourcing platform that is part of Zooniverse. Provides a variety of projects for transcription, along with blogs and other sources of information to help participants understand the context and use of collections in research

- BIOSPEX (<http://biospex.org>) -- a tool to provision, advertise, and lead public biodiversity digitization expeditions.
- Worldwide Engagement for Digitizing Biocollections: WeDigBio (2014–present, www.wedigbio.org) is a citizen science event of collaborative, interactive biocollection transcription with strong formal and informal educational goals. As a global event, WeDigBio encourages participants to learn about biodiversity science locally and internationally. Several university classes participated in the inaugural WeDigBio event (2015) and used formal biodiversity-related activities such as a lesson plan developed jointly by collectionseducation.org and WeDigBio, and lessons created for specific class use.
- The Smithsonian Transcription Center has developed best practices for transcription projects: <https://transcription.si.edu/tips>
- Digivol (<https://volunteer.ala.org.au>) is a crowdsourcing platform developed by the Australia Museum that is used by 32 institutions

NOTE: There is substantial effort being leveraged on these goals by numerous players. However, integration of the disparate initiatives and entities into a collaborative, consistent framework is still needed. This integration must align the different components into a data pipeline that encompasses all of the major players in this realm—field biologists, collections, and collections managers, aggregators, researchers and other end users, publishers, data packagers (like Genbank, and so forth) and funders (like NSF). The goal for some sort of governance, or national coordination function, remains unfinished, which contributes to the current lack of integration, despite the great deal of progress that has been made. Forward movement in education, formal, informal, and career, is also evident. However, full implementation of education initiatives will have to be integrated, in an iterative process, with the ongoing developments in the data systems and knowledge bases that support NIBA goals.

Restatement of Goal:

Infuse specimen-based data into all levels of formal and informal education resulting in a diverse and empowered data literate society, data-enabled workforce, and cross-disciplinary data-savvy researchers that leverage the vast collection infrastructure, meta-data, resources and information as we address the grand challenges of science in the 21st century.

Proposal for the Future:

- Integrate educational efforts and centralize products to efficiently use time and resources.
- incorporate biology educators into research efforts.
- Develop meaningful undergraduate research experiences incorporating specimen curation continuum.
- Provide best practices for mentoring student workers in natural history collections.
- Develop and refine educational modules involving natural history collection data:
 - In introductory coursework
 - Across the biology curriculum
 - Across the science curriculum
 - Across all disciplines
 - Disseminate materials

- Introduce modules in educational venues (education conferences and journals)
- Incorporate partners into digitization efforts
- Better prepare educators to use biodiversity collections data.

Citizen Science

- Develop materials to incorporate citizens as biodiversity sensors.
- Incorporate students as data providers.
- National Network of CitSci Endeavors—integrate efforts.
- Link out to other citizen science initiatives (NPN, National Parks).
- Regional networks.
- Work with eBird, eButterfly, ePlants.
- Crowdsourcing.
- Best practices for quality control.

Immediate Action Items:

- No immediate actions were specified.

General Comments on Report

BCoN thanks everyone who provided comments and suggestions on drafts of this report. We have endeavored to incorporate clarifying changes where possible. Other comments are more significant and worthy of attention in future workshops and programs, and thus have been summarized below. We have attempted to remove personal information from comments.

Comments, Questions and Suggestions from the Community

Thanks for all the efforts that went into this valuable program & document. It has been very informational just to read, and I want to mostly endorse goal #3. The mentorship program & certification program should especially be pursued.

BCoN Response: Thank you, we are happy to learn it was informative and will continue to explore the issues raised.

In reference to: "Goal 4.5: Expand NIBA to include digitization of federal collections; NIBA and NSF outreach to federal interagency committees and agencies on leveraging NIBA resources. Nothing was reported."

Several federal institutions are working in collaboration with current TCNs including the Fossil Insect TCN (Smithsonian National Museum of Natural History and Florissant Fossil Beds/National Park Service) and the EPICC TCN (Smithsonian National Museum of Natural History). These institutions are funding the digitization through internal federal sources.

BCoN Response: Thank you for drawing this to our attention.

This is a valuable effort, but I remain concerned that the focus on digitization and database construction hasn't been accompanied by much support for the tasks of data generation and verification. Taxonomic expertise needs to be funded and databases need to have mechanisms to receive and process feedback for corrections. Without attention to data quality, all the databases merely make it easier to generate unjustified analyses.

BCoN Response: Thank you for raising this important issue. These topics are likely to be considered at workshops being planned for 2018.

One issue I do not see addressed at all as part of Goal #2 is cybersecurity.

BCoN Response: Thank you for raising this important issue.

The reality of climate change and its effect on biodiversity and biomigration is barely mentioned. There is an assumption that floras and faunas will remain static, and analysis of data will largely focus on climate-related minor changes in situ.

Given that a recent UN report asserts that warming will easily reach three degrees Centigrade, biotic displacement on a large scale is inevitable, guided and constrained by those corridors that are available. Whatever features are envisaged for analysis of digitized data, they should include presentation of the data that may be monitored and predicted as to changes in biotic communities and ecosystems through time across space. The characterization of communities as represented in present collections data should be assumed as only a benchmark for rapid and major change.

One can envision that it is a certainty, given present policies, that rapid displacement over a very few decades of both natural biota and of humans and their works, will become a major concern at the highest levels of government, and that monitoring and prediction of such displacement will become critically important. Planning for this should be part of the Implementation Goals of NIBA.

BCoN Response: Thank you for drawing attention to these important issues. These topics will likely be included in discussions being planned for later in 2018.

Goal 2.3

If this activity could be automated (e.g., Genbank identify all sequences tied to vouchers at an institution and send them to the collections), this activity could be significantly enhanced.

As we learned at the recent iDigBio summit, 72% of TCNs are using Symbiota as their collaborative interface. This includes ALL of those involving herbaria where duplicate matching can be used quite effectively. An ongoing issue in the herbaria community is trying to use Specify collaboratively!

Goal 2.4

Interoperability: this is important issue that should be considered soon.

BCoN Response: We agree and it is the subject of a February 2018 workshop that will be convened at the University of Kansas.

Goal 2.6

Are individual organizations able to join these institutions, at what cost?

Goal 2.7

One problem that might hinder usage of these resources; professionals may not trust the results. I fall into that "boat" - I don't use the TNRS, preferring to investigate names on Tropicos and IPNI.

Immediate Actions, Goal 2

Excellent idea! I can think of two gaps:

- One that several herbaria participating in TCNs have discovered (sometimes the hard way) - using Specify in a collaborative manner (being able to take advantage of duplicate matching, OCRing of labels, etc.) is not yet practical. Interchange of data between the two programs is not a two-way street.

2 - are there any plans to integrate data from Plants.jstor.org into iDigBio?

Goal 3, Proposal for the Future

As I read this section, I am struck by the presence of the word "curator" and the absence of "collection manager". In some institutions, the "curator" is far less "hands-on" than in the past - and often they are significantly removed from the curatorial aspects of data management. I would prefer to see both titles included.

BCoN Response: Great point, we'll endeavor to do this in the future.

Goal 5.1:

Are the "results" of that Summit going to be publicized?

BCoN Response: Yes

Goal 6: Immediate Action Item

I would suggest one here - efforts toward making it easier to integrate the results of crowdsourcing into other data management systems (e.g., Specify). I know that question keeps some from participating in ANY crowdsourcing....

For more information about this report or the Biodiversity Collections Network, please contact Robert Gropp at rgropp@aibs.org or 202-628-1500 x 250.