Extending U.S. Biodiversity Collections to Address National Challenges

Draft for public comment, 14 Jan 2019

I. Executive Summary

A workshop held 30 October - 1 November 2018 at Oak Spring Garden in Upperville, VA under the leadership of the Biodiversity Collections Network (BCoN) addressed the future deployment of data held in U.S. biodiversity collections for research, policy and education. In their discussions, participants drew heavily on recent literature as well as surveys, meetings and workshops held over the past year with the stakeholder community of collections professionals, researchers, and educators.

Arising from these deliberations is a consensus to focus future biodiversity documentation on building a network of extended specimens that represent the depth and breadth of biodiversity held in U.S. collections institutions. The extended specimen will consist of the physical voucher and any preparations (e.g. tissue samples); digitized representations such as occurrence records and images; derivative products such as gene sequences or metagenomes; and taxon- or locality-specific data such as observations, phylogenies or species distributions. We will extend existing specimens by data linkage through unique identifiers, taxon name and location. New collections, needed now more than ever to inform solutions to societal problems, should be "born-extended", i.e., accessioned with a full suite of supplemental data.

Collectively, these extended specimens will form a network of linked data to enable exploration across taxonomic, temporal and spatial scales. Such exploration will help us understand the rules that govern how organisms grow, diversify and interact with one another, and how environmental change and human activities may affect those rules. As a resource for formal and informal education (including citizen science), the extended specimen network will afford scalable learning opportunities for K-Life in data literacy as well as biological science and the humanities.

To create this resource will require continued specimen digitization, new collections, standardization of existing digital data to facilitate discovery, and implementation of a robust specimen identifier tracking system. It will also require new approaches to data sharing and collaboration, partnerships with national and international data providers, computer and data scientists and educators.

This initiative requires episodic funding for collection of new specimens as well as for digitization and curation of existing resources. It requires long term funding for a central organizing unit with responsibility for community coordination, education, mobilization, and maintenance of the central data repository and the network infrastructure. Representatives of the central organizing unit, collections institutions, and professional societies should oversee the implementation of the agenda.

II. Background

In 2010, the U.S. National Science Foundation (NSF) convened several meetings of researchers and collections professionals to consider a national plan for the digitization of biodiversity collections. The result was the [Network Integrated Biocollections Alliance (NIBA) strategic plan], which laid out the need for the digitization of collections and proposed a structure for the effort. The plan called for the creation of thematic collection networks that would digitize specimens to create a dataset for addressing particular research questions, as well as for a central hub that would organize the effort and provide training and support for the networks. NSF responded to this NIBA plan by creating the Advancing Digitization of Biodiversity Collections (ADBC) program, a ten year, $100 million commitment that has made annual awards since 2011. Several years later, representatives of the biodiversity collections community crafted the NIBA Implementation plan, which established six implementation goals. The Biodiversity Collections Network (BCoN), a five year Research Coordination Network (RCN) NSF award (DBI-1441785), was funded in 2014 to bring members of the community together to address collections needs that fell outside the scope of ADBC and its strict digitization mandate.

Since the release of the NIBA plan, we have made transformational progress in the digitization of specimens and sharing of specimen data. The ADBC program has funded 23 Thematic Collection Networks (TCNs) and 29 Partners to Existing Networks (PENs) to add additional collections to the network. To date, the TCNs have collectively digitized 62 million specimens from 915 collections held in 317 institutions and have provided training and work experience for thousands of students and emerging professionals. The ADBC program has provided continual support to University of Florida and Florida State University for hosting Integrated Digitized BioCollections (iDigBio; DBI-1115210; DBI-1547229), the central coordinating unit for the digitization effort. iDigBio provides training in digitization and data mobilization to participating institutions, and shares digitized data through its iDigBio Portal, which contains 115 million specimen records and 27 million associated media records from 1,576 datasets. iDigBio has also carried out a vigorous outreach program, with symposia, webinars and workshops on digitization methods, data manipulation and sharing, and use of digitized data in research. They have also developed and disseminated a vast array of best practices and standards associated with digitization through their website. There is a growing body of research using data digitized served through iDigBio and other ADBC-sponsored portals – more than 500 citations in 2018 so far (see https://www.idigbio.org/research).

The end of the ADBC funding program is on the horizon. Grants funded in 2021, the last year of competition, will conclude in 2023 or 2024. iDigBio will continue to function through the end of the last ADBC grant. The processes, tools, and protocols developed provide a foundation from which innovative new research fields and solutions to national problems may be identified.

This report summarizes the results of a workshop held 30 October - 1 November 2018 at Oak Spring Garden in Upperville, VA. BCoN initiated the workshop, which was co-hosted by the Oak Spring Garden Foundation. A list of participants is included in Appendix A. The goal of the workshop was to identify a national strategy or agenda for the next phase of the effort to deploy the data held in U.S. natural history collections for research, policy and education. In consideration of this topic, we reviewed the 2017 document benchmarking accomplishments of the past decade against the NIBA goals, a modified version of which comprises Appendix B. Participants also reviewed responses to surveys and discussion sessions with the primary stakeholder community of collections professionals, researchers and educators. The outreach efforts are summarized in Appendix C, and citations for literature reviewed in preparation for the workshops can be found in Appendix D.

III. The Extended Specimen Network

Building on the accomplishments of the past decade, we propose to transform our rich heritage of collections and associated data into a powerful new source of knowledge to address national priorities. This resource will have transformative potential to address long-standing questions about the breadth and complexity of biodiversity. Central to this agenda is the creation of extended specimens that represent the depth and breadth of biodiversity held in U.S. collections.

The concept of the extended specimen, introduced by Webster (2017), elevates and expands the physical specimen with an augmented digitized specimen record by associating data types including field recordings, computable and semantically rich descriptive content, and internet-scale connected data resources that support discoverability of this content in novel ways at multiple scales.

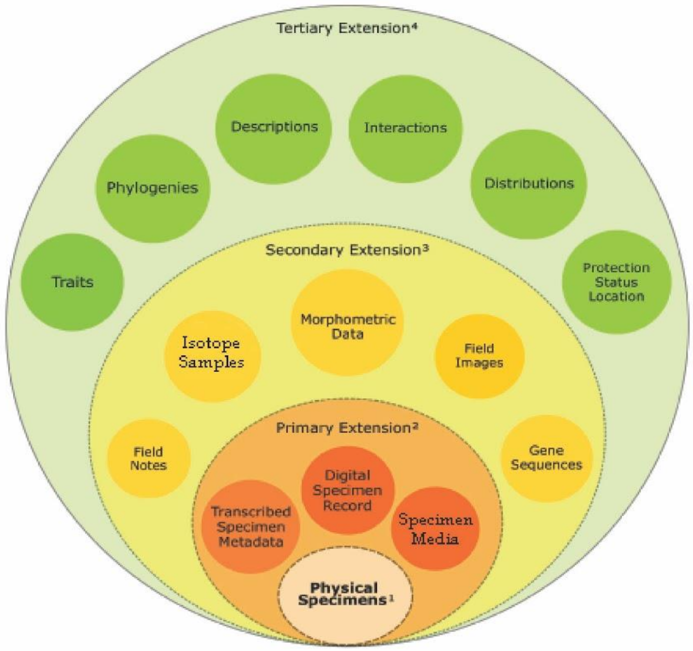

Figure 1. *Components of the extended specimen. [1]The approximately 1 billion physical specimens that reside in the approximately 1600 museum collections in US, including derivative collections, such as tissue samples. [2] Data and images, audio and video files held by collections that hold physical specimens and also contributed to the national specimen database. [3] Data that reside in private collections, institutional repositories, and national/international repositories, often disconnected physically and digitally from the original specimen, which may link to individual specimens. [4]Taxon or occurrence locations level data linked to more than one specimen record.*

Enabling Scientific Discovery

Specimen digitization provides critical information on species occurrence and co-occurrence over space and time, enhances species discovery, and provides a means to scale up and reduce uncertainty when examining overall patterns of diversity. Recognizing the value of this resource, members of the stakeholder community identified increasing its use in research as the top priority for the next phase of data mobilization from collections. Indeed, a significant body of new collections research has emerged in the past decade. The bibliography compiled in preparation for the Oak Springs workshop (Appendix D) includes citations of articles on the use of specimens in research in climatology, genomics, human health, and phenology and trait characterization.

Combined with data integration technologies and relevant data layers, extended specimens will become part of a powerful data network. By anchoring this network in the physical specimen, which places an organism at a particular place and time, we can link data spanning the ecosystem to the genome and microbiome to understand better the rules that structure biodiversity. Specimen and habitat images, video, audio recordings, and other media connected to physical specimens can reveal undiscovered traits, mutualistic or parasitic associations, and also shed new light on previously studied phenotypes. When applied to museum specimens, techniques such as CT scanning allow researchers to view internal structures in 3D, allowing visualization of previously hidden-suites of characteristics, allowing for novel comparisons across diverse groups of organisms.

Perhaps the most exciting new research area to which extended specimens can contribute is the exploration of the traits that compose an organism, e.g., internal and external structures, physiological processes and behavior, and interactions with other organisms, their environment, genotype and phenotype. A better understanding of traits has direct benefits to our society and quality of human life, including understanding of how zoonotic diseases are transmitted (and controlled), how crops can be more effectively and efficiently grown in changing climates, and how we can sustain and use biological resources in our oceans. We need to develop knowledge graphs (ontologies) to make trait data gathered from specimens interoperable, which requires new infrastructure and computer science expertise (Kissling, et al. 2018). The result is an integrated global enterprise of specimen-based data that will benefit non-biodiversity related research and commercial communities, creating the ability to engage, communicate, and exchange data effectively and efficiently. The extended specimen network will allow for new avenues of investigation, expedite existing investigation, and provide a deeper resource for making science-based policy decisions.

The extended specimen network will also facilitate the still incomplete work of documenting and naming the organisms that make up earth's biodiversity. We will achieve this documentation by new exploration in tandem with review of existing collections. Estimates suggest that 75% of undescribed species in the world are already represented in collections, but are misidentified or languish in an unidentified backlog (Bebber, 2010, Kemp, 2015). Incorporation of machine learning and other data science and engineering techniques, including pattern-matching techniques that can use the full range of data available about specimens of known species, can help to find hidden novelties from both images and specimen data records, including geospatial and genomics-linked resources (McAllister et al., 2018).

The extended specimen network will be a context-rich source of species occurrence data that can inform land use management and conservation policy in a more integrative manner than is currently possible. A user today wishing to investigate the biodiversity in a particular region can almost certainly retrieve a set of relevant occurrence records from U.S. collections through one or more data aggregators. However, use of these data will be more efficient and powerful when the occurrence data link to other information sources to allow them to ask such questions as: Does the available data comprise a representative set, or are critical data lacking because key specimens have not been collected or digitized? How many different species, as opposed to different species names, occur in a given region, and are there common phylogenetic trends in the species assemblage? Do the organisms of the region occur in populations that are genetically distinct from other populations of the same species? Have unique interactions among organisms been documented in the region of interest?

Empowering 21st Century Learners through Biodiversity Knowledge

The collections community has made notable strides to promote their use in formal and informal education (Appendix B). We can now develop and implement focused and inclusive education strategies and materials that engage a richer array of learners and result in improved educational outcomes. We can also engage educators who in turn will engage new student audiences and novel stakeholders. To train the biologists of the future and improve science literacy in general, we must make it a priority to infuse specimen-based learning into all levels of formal and informal education.

We define education broadly in this report. For our purposes, *Formal Education* encompasses K-12, undergraduate, and graduate "general biology" formal education. *Advanced Professional Training* applies to emerging biodiversity researchers in graduate programs and working as post-doctoral researchers. This category also includes existing scientists, curators and museum professionals with skill gaps. We use the term *Informal Education* as a broad category that spans everything from traditional museum experiences to rapidly developing citizen science initiatives with incredible potential to engage a broad audience in gathering and transcribing biodiversity data.

Formal Education. Natural history specimen-based data can enable training and education of data savvy scientists and engaged biodiversity enthusiasts. Specimen-based data make science accessible through the specimen itself, which is tangible, place-based, and interesting, as well as through aggregated specimen data that are verifiable, relevant, and a logical gateway to data literacy (Cook et al., 2014, Powers et al., 2014. Monfils et al. 2017). Biodiversity data usage aligns naturally with core content common to all K-12 science curricula, including evolution, biodiversity, systematics, taxonomy, and ecology. These data integrate into the curriculum without a dramatic increase in the content to be covered. Data literacy skills can be introduced within the context of exploring topics including climate change, spread of disease, species conservation, interspecific interactions, and invasive species. Students can use specimen-based data to generate their own hypotheses and test them using the biodiversity data available online. Students collecting their own specimens can engage in the full data lifecycle from documenting data collection (including metadata), to prepping and archiving specimens, recording specimen data and data standards, and submitting data to a formal data archive.

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. Materials developed or endorsed by BLUE support the training of diverse, competent, and engaged young biologists who are well prepared for a broad set of career paths generating and utilizing biodiversity data to address scientific issues of critical national and global importance.

As we look to novel uses of specimens in research, we need to avail ourselves of the highly diverse millennial workforce. Diversity in research teams provides new perspectives and ideas leading to novel solutions and increased productivity (AlShebli et al., 2018). Increasing the presence of historically underrepresented groups within the biodiversity sciences requires a concerted community-wide effort of those actively involved in these fields. The place-based capacity of collections data combined with the social and societal relevance of biodiversity science can serve a role in creating inclusive, culturally relevant, and socially conscious educational materials that engage a broad and diverse audience in biodiversity science. When creating educational materials or speaking on the value of biodiversity data, we must document and share the history of the contributions of diverse scientists to natural history and engage a diversity of natural history researchers and collections professionals as spokespeople for the field of biodiversity science. We need to engage indigenous communities to capture and preserve knowledge and perspectives, and we need to be working internationally to create programs with a universal design. Providing effective rationales for pursuing biodiversity related professions, clear models of career paths within these professions, outlines of the available educational opportunities leading to these professions, and sustained support as students move through the educational system are critical for diversifying and empowering the emerging workforce.

The biological collections community can engage a larger set of stakeholders and extend the reach of the biodiversity data beyond the biology classroom by providing the datasets, experiences, and materials that engage data-centric non-biological sciences, bridge the liberal arts disciplines, and include the arts communities to achieve the goals of STEAM education. As within the biology classroom, we need to be conscious of the different resource needs of prospective students and educators as we create inclusive materials that meet those requirements and provide individualized support for successful implementation.

More than any time in the past, graduate students in research-driven Master's and Ph.D. programs can gain diverse and practical skill sets via the analysis of biodiversity data, including those related to taxonomy and natural history, (paleo)genomics, statistics, machine learning, and computer science. Large databases that the broader natural history collections community have been developing, maintaining, and extending in recent decades perfectly suit graduate training that incorporates data analysis, including statistical analysis, machine learning, and artificial intelligence. Further, the trained students may go on to mentor the next-generation of natural history collections-based scientists, or assume careers outside of our core biodiversity science community (e.g., as data analysts in biotech, agriculture, or other industry), where they will bring awareness to the importance and promise of natural history collections.

With global issues of climate change, invasive species, zoonotic diseases and food availability and security, we need to train the next generation to think creatively, be open to new ideas and new technologies, and work with others to generate ideas and problem solve. Biodiversity scientists will require advanced training in a diversity of fields (e.g., integrative and comparative biology, systems biology, evolution, ecology, developmental biology, computational biology). Soft skills, including flexibility, communication, management, responsiveness, and the ability to cooperate with others, are fundamental to success. Students will need to work with multiple mentors with diverse expertise in order to train for addressing large-scale multidisciplinary science questions, and thus we need transformative programs to affect this paradigm shift in graduate education.

The NSF Postdoctoral Research Fellowship in Biology Track 2: Research Using Biological Collections (2015–present) is an excellent example of the type of advanced training programs needed to advance this agenda. This program funds postdoctoral researchers who perform collections-based research, often in novel or cutting-edge ways. Importantly, by explicitly funding novel uses of collections, this program has resulted in a variety of uses of collections that a "typical" museum researcher (i.e., phylogeneticist, taxonomist) would not necessarily envision, including evaluation of plant responses to rising CO_2 availability using herbarium collections, understanding the evolution and development of tetrapod olfaction through histology and imaging of museum specimens, and testing hypotheses of invasive species establishment using trait data from digitized, curated ants. Such programs, including former cost-effective research grants like the NSF-DDIG program, are important for training future curators and other investigators.

Advanced Professional Training. Collections community survey respondents cited the lack of programming or other data science expertise as one of the greatest obstacles to the digitization of their specimens and the use of resulting data in collections management and research. Therefore, we also need expanded training programs to prepare the current and next generation of collections managers and curators to digitize, augment and maintain large data sets.

The collections community will need museum studies programs that integrate fundamental data and data management skills as well as the basics of coding. With the rapidly changing requirements of collections professionals and the considerable skill set associated with digitization, aggregation, and management of large biodiversity data sets, we need to consider training that can meet the needs of professionals at different stages in their career. Since 2011, iDigBio has been highly effective at creating inclusive and widely applicable training workshops and webinars. These opportunities to learn new skills have helped to create a broadly inclusive collections community as well as improved understanding between data managers and data users

(Seltmann, et al. 2018). The Carpentries have been useful in training the research community in coding and foundational data science, but the need far exceeds the supply and the specific challenges of museum data have not yet been written into their curricula. Integrating continuing education with ongoing formal education, drawing on the skill sets of science education researchers, creating standard materials and gateway workshops to teach about different aspects of collections management and biodiversity science are all critical in leveraging our past successes to create a sustainable knowledgebase.

Informal Education. An increasing number of citizen science projects and focused programs at, or hosted by, museums and natural history collections are engaging the public in biodiversity science, and as our digital resources continue to expand, so too will informal education opportunities. Indeed, we probably cannot obtain the full expression of the extended specimen network envisioned in this agenda probably without strong involvement of the citizen science community.

A number of citizen science projects are based on monitoring biodiversity. For example, eBird, eButterfly, iNaturalist, and US National Phenology Network, to name just a few, provide platforms for minimally-trained individuals to contribute sightings or recordings of organisms or a particular attribute of an organism, e.g., its phenological state. For individuals interested in a deeper understanding of biodiversity, offerings such as Master Naturalist programs can provide the opportunity for members of the public to become experts in their local flora and fauna. Biodiversity monitoring and naturalist programs led by museum and collections staff link professional scientists with the public and provide an authentic, interactive educational experience.

The Urban Nature Research Center at the Natural History Museum of Los Angeles County is a strong example of what can be done when museum professionals collaborate with citizen scientists to investigate urban biodiversity. Notably, UNRC researchers and participants in the citizen science projects they have developed have discovered numerous species new to science (e.g., Hartop et al., 2016), documented range expansions of introduced species (e.g., Vendetti et al., 2018), and have demonstrated natural history museum-based citizen science contributions to species conservation (Ballard et al., 2016). Specimens remain the gold standard for verifiable data, yet as UNRC researchers and collaborators find (Spear et al., 2017), citizen science observations with specimen data are perhaps the most effective and efficient way to scale up data collection to address many of today's most pressing biodiversity challenges. In addition to the direct research implications of these citizen science projects, participants also often help provide specimens to collections. In doing so, citizen scientists learn collections skills while improving the representation of recent records within our collections.

Internet-based projects can involve the public directly in contributing to collections-based science and databases. Projects hosted by Notes From Nature, Smithsonian Transcription Center, and CitSciScribe that invite the online public to add digital data to images of specimens. Tasks range from transcription, to morphological measurements, to phenological annotation, and the majority relate directly to an active research project. While tasks are online, researchers and scientists actively communicate with participants to answer questions and ensure that volunteers complete tasks as designed. WeDigBio, an annual global transcription event, has united these efforts to work collectively towards digitization goals over the course of four days each year. Such programs engage participants from a wide range of ages, abilities, and interests, and with minimal start-up costs. This approach creates an inclusive and diverse group of individuals working to advance biodiversity science.

Additionally, the success of small scale efforts by institutions to engage recent immigrants, developmentally challenged, and previously incarcerated citizens in efforts to improve language, keyboarding and workplace skills indicates work in collections can improve quality of life in addition to serving as an entry point for new scientists (*Moving the Needle: Broadening Participation in the Biodiversity Sciences* webinar series). Many students engage in citizen science platforms as part of formal education activities, creating a point of intersection between formal and informal education that can be leveraged for addressing the needs in both communities.

By design, citizen science projects turn to established best practices and procedures for quality assurance and control. Such practices include training documents and videos, within-project quality checks, and post hoc assessments. Citizen science data standards and best practices have been created and vetted by members of international organizations such as the Data and Metadata Working Group with the US Citizen Science Association and the Citizen Science Interest Group with the Biodiversity Information Standards (TDWG) organization. These measures ensure that the work completed by citizen scientists is research-ready. However, we must continue to develop and improve training materials, best practices and ongoing support from the scientific community to maximize the impact of citizens as real-time biodiversity sensors.

Enable Seamless Data Integration, Attribution and Use Tracking

The key to assembling extended specimens and linking them into a robust data network is the standardization of existing digital data and implementation of a universal specimen identifier tracking system. Data integration is required at multiple levels - at the specimen level to link individual preparations (e.g., vouchers, tissues, gut contents), at the database level to integrate taxonomic interactions (e.g., predator:prey, host:parasite, plant:pollinator) and beyond collections to link other sources of relevant data (e.g., observation data, environmental data, geographic data, ecological data, satellite data). In addition to permitting the linkage of the digital specimen record to related data sources, a tracking system will allow researchers to attribute data usage in a standardized manner and will permit compliance with new reporting requirements for the acquisition and use of biodiversity data.

New technologies, such as blockchain (https://en.wikipedia.org/wiki/Blockchain, Halamka et al., 2017, Hilsberg, et al., 2018), may provide viable solutions for how data and products can be stored, linked and shared with collaborators and stakeholders, facilitating transparency and traceability while maintaining data integrity, standardized auditing, and formalized contracts for data access. This decentralized approach to data exchange has already gained significant traction in the corporate, academic, and federal arenas. Some discussions of how museums can utilize blockchain include ways to preserve data in a low-cost manner while enhancing the reachability of specimen-based data to students, researchers, teachers, etc. as well as for enhancing collections management (Campbell, 2017, Lemle, 2018)

Attribution measures are an important advocacy tool for collections to demonstrate their value and advocate for collections management resources (e.g., staff, funding,). Collections are generally unable to showcase their full contributions to specimen-based research through citation in publication, vouchering of Genbank sequences, products created from direct specimen use (images, CT scans etc.) or use of data downloads from data aggregators due to limited connectivity between the collections and research, aggregator and publication communities. The overarching reason for this is the lack of a reliable mechanism of tracking such use. There is also a need for attribution metrics for the individuals involved in data curation and collections management activities (collectors, catalogers, determiners, georeferencers, data users and augmenters) to allow for individual acknowledgement and assessment of the work done in providing or augmenting digitized records.

The proliferation of data aggregators makes it increasingly difficult for collections to track use of their data effectively by the ever increasingly varied end user communities now using biodiversity data. The duplication of both mechanisms for publishing records and data manipulation tasks has created an environment of uncertainty and incompatible versions of individual records. The necessary data integration and attribution infrastructure requires both technological and social solutions -- technological solutions to facilitate the linking of disparate data elements, and social advocacy to ensure that all those involved in creating and using the data pipeline are aware of the need and benefits of repatriating products of use back to collections. The Global Biodiversity Information Facility (GBIF) is attempting to address issues of attribution and integration by proposing that the major aggregators work towards unifying their disparate data caches and data tasks operating on that data stream (taxonomic authority, geographic authority, Darwin Core standards, etc.) to simplify the data aggregator landscape and bring about data consistency.

A consequence of the greater discoverability of biodiversity collections through digitization is new requirements for the acquisition and use of these data. The Nagoya Protocol is a supplementary agreement to the Convention of Biological Diversity (CBD), that establishes an international legal framework for access and benefit sharing (ABS) of genetic resources. It requires that countries providing specimens define their access procedures, calls for users to share benefits upon their use, and establishes an international framework to ensure compliance with domestic access and benefit-sharing laws. By necessity, this new regulatory landscape will affect how biodiversity collections are managed and used, and since many of these issues are beyond the traditional scope and expertise of the current biological collections community, it will require an interdisciplinary approach to the development of technological solutions and training.

Improved specimen tracking will not only facilitate global compliance with the emerging policy and legal issues such as the Nagoya Protocol, but also will support better metrics for data use and create the potential for cost-recovery for use when specimens are used in commercial enterprise. The ability to work with novel research (e.g. pharmacology, human health, food security) and commercial communities (e.g., pharmaceuticals, agriculture) will demonstrate the value of collections outside the immediate stakeholder community and contribute to the sustainability of biodiversity collections.

IV. Building the Extended Specimen Network

Fill Gaps in Biodiversity Data

Biological collections comprise the most comprehensive record of life on our planet, yet their potential cannot be fully realized until the data contained within them are revealed. Although institutions maintaining biological resources have a tremendous opportunity to leverage their specimen holdings and expertise to address grand challenges in science and society, many of the approximately 1,469 U.S. collections included in the Collections Catalog created by iDigBio do not have accurate estimates of their holdings for major taxonomic groups. Many types of collections are still underrepresented in existing collection indices (e.g., small, regional and single-investigator collections) and the location of ancillary specimens that may originate from other taxonomic disciplines and are disassociated from related material (e.g., parasites, commensals, host, gut contents, substrate and water sample collections) is often undocumented. The sheer volume and varied storage of material for certain groups of organisms, and the lack of taxonomic expertise, resources, and collective standards of assessment, often prohibit a more complete characterization of national holdings, in even the most general terms. Additionally, many species remain unknown, and the groups that are taxonomically most poorly understood often represent the groups for which the majority of existing collections are undigitized.

Comparison of digitization progress in different taxonomic groups reveal stark disparities. Through ADBC funding, essentially all collections with a mycological component have digitized at least some of their fungal specimens. An estimated 43% of all fish collections

(60 out of 143) share some digitized data through an aggregator; iDigBio holds data for 38 of these collections, including 3,107,368 specimen records (typically referred to in institutions as "lots") with 44,128,165 specimens (Singer, et al., 2018). Rough estimation based on data from the iDigBio data portal, various herbarium consortium networks and *Index Herbariorum* reveal that at least 437 of the 659 herbaria in the U.S. (66%) have digitized some of their holdings, resulting in at least partial digital records for approximately 26 million of the national total of 76 million herbarium specimens, (34%). In contrast, estimates indicate that only 6% of all insect collections have been digitized (Cobb, et al., in preparation). A recent article on the EPICC TCN (Eastern Pacific Invertebrate Communities of the Cenozoic), reports, "Our group...quantified just how much "dark data" are present in our joint collections. We found that our 10 museums contain fossils from 23 times the number of collection sites in California, Oregon and Washington than are currently documented in a leading online electronic database of the paleontological scientific literature, the Paleobiology Database." (Marshall 2018 and Marshall et al., 2018)

A lack of complete digitization should not prevent the community from projecting future needs. A combination of existing digitized records coupled with collection-wide estimates of holdings can help provide a basis for understanding gaps in our collections and associated biodiversity data. These types of assessments would allow us to determine how many specimens we need to address a range of global change issues, including questions that have not even been formulated yet. Discipline-specific efforts such as Index Herbariorum, the ASIH Standard Symbolic Codes, iDigBio's Collections Catalog, GBIF's provider list, and NCBI's institution and collection tables provide a partial summary of collections throughout the world but this inventory is far from complete, both in scope and richness of individual record data. Endeavors to formally codify the metadata fields necessary to describe collections and institutions have begun through the TDWG Natural Collections Descriptors (NCD) working group. However, these descriptors are not implemented yet, and a renewed effort may be required to gain a complete assessment of collections and their holdings in the US and beyond.

Although we have made great progress to date, we need to digitize many more extant and fossil specimens, as well as continue collecting new specimens to document and maintain an accurate representation of changes in species diversity and distributions over time. Species rich, yet understudied groups of organisms will require new collection campaigns to fill taxonomic gaps. We will also need to address the issue of geographically and temporally biased sampling of taxa– even groups that are taxonomically well understood and digitized may not be well sampled in space and time. If biodiversity specimens are to be a resource for documenting environmental change, then researchers must continue to collect them into the future. Importantly, a more holistic, next-generation approach to the collection of new biodiversity specimens must evolve. As noted by Schindel and Cook (2018), standards for collection and preservation of natural history specimens have changed relatively little over centuries, with the exception of recording the global positioning system (GPS) location and preserving new types of samples for subsequent DNA and RNA analysis (e.g., tissues, gut contents, parasites). A next-gen approach could focus on nested sampling that extends beyond the single organism (e.g. a single plant), to its biotic associates (e.g., soil microbes, epiphytes, endophytes, and parasites spanning from viruses to insects and fungi) and its environment (e.g., community composition, microclimate, macroclimate, and habitat quality).

We must explore new strategies for how to reach a critical mass of digitized specimen data. For collections of some of the larger taxonomic groups, rapid digitization using skeletal records with minimal fields of information or lot-based digitization, as well as technological solutions (e.g. robotic imaging stations for specimens in drawers, lot imaging) have been, and will continue to be developed. These records can act as proxies for the undigitized components of biological collections, with specific emphasis placed on the data types that augment primary occurrence data (i.e., observation data, images, field notes, movement data, slide collections, vocalizations, video, DNA sequences). This call to action to expand collections and their access via digitization will also require mobilizing new researchers and personnel to enact project-focused collections that are digitized upon acquisition (e.g., Contreras 2018), increasing person-power applied to digitizing existing collections.

Incorporating biodiversity data effectively into research and other initiatives will depend on data accessibility through a permanent central portal, as well as continued integration of new and existing data across the community. In the realm of collections, this includes the seamless connection of specimens and derived data at all levels. Specimen preparations (e.g., tissues, skeletal elements, gut contents) should be linked back to the original voucher specimen within collections in order to associate all products of collections research. The integration between different collection types (e.g., predator-prey, host-parasite, tissue-voucher) and datasets outside the collections-based research realm ensures the future utility of collections to a rapidly expanding population of end-users. Over the past ten years, we have worked to accumulate massive molecular, trait, and environmental datasets that we can now be couple with digitized specimen data and use in ways not previously imagined. When combined, these resources will enable new discoveries and facilitate computational connections between these different data types, unleashing inquiry-based science in novel, and perhaps even unexpected, ways.

The National Ecological Observatory Network (NEON) collects and archives approximately 80,000 samples annually through its continental-scale environmental monitoring platform. Each sample receives appropriate physical and digital curation to enhance sample discoverability and enable transformative research. NEON facilitates the extended specimen paradigm by ensuring that all archived specimens are associated with precise spatial and temporal information from their collection along with any available genetic or taxonomic metadata. Depending on the data product and sample type, NEON provides specimen images, morphometric data, genetic sequences and taxonomic data. The NEON Biorepository uses a Symbiota platform for its specimen portal. Researchers will have access to NEON samples to conduct additional analyses; results of these analyses will be linked to the samples themselves on the portal.

Complete and Improve Existing Digitized Data

The effort on the part of U.S. collection holding institutions to generate 62 million digitized specimens during the past decade has been truly remarkable, and the value of this resource has wide-ranging implications for research, education and collection management. However, a large proportion of these newly digitized specimen records can be improved and augmented. Some lack locality, date and collector information, and geocoordinates have not been provided for most. The incomplete transcription and georeferencing of specimens inhibits the comprehensive use of these data, and we must complete this work in order to maximize their value in the extended specimen network. Development of tools that can help fill in, or at least infer, missing values must be a high priority, perhaps taking advantage of a combination of image analysis (including optical character recognition) and data pattern matching.

The structure of specimen occurrence metadata across biodiversity collections is standard enough to have allowed its definition in the Darwin Core schema. However, the data values, gathered across vast temporal, geographic and taxonomic ranges are usually not standard even within a collection. Momentum is gaining in this area with the implementation of standardized data quality test and assertion tools by the major aggregators and driven by work done by the TWDG Data Quality committee. The stakeholder community considers the inability to query databases reliably using key metadata criteria to be one of the greatest obstacles to greater use of specimen data in research and education (Appendix C). Because the use of specimens and associated data by the research community is predicated on the accessibility and accuracy of those data, more work is required to improve data standards, increase access to taxonomic authority sources, and ensure adoption of standard vocabularies and data models for published data. Further improvement may necessitate the development of additional tools that build upon the technical groundwork and data pipeline construction completed during the first round of ADBC funding.

Inspire an Open, Integrated Collections Alliance

To realize the full potential of digitized biodiversity collection data, a more community-minded approach to data gathering and sharing must evolve, and the training and reward structures for collection community professionals must be adjusted to place higher value on collaborative activities. Such changes will foster an agile community capable of supporting curation, research, and the mobilization and use of data by a much broader array of scientific and commercial interests. We must train collectors of new specimens to approach their collecting in a way that maximizes efficiency and accuracy of downstream digitization and analysis to prevent continued accumulation of undigitized and unprocessed material accruing within collections.

Collections must insist that data from new collections be structured according to community standards and that data flow seamlessly into the institutional, national, and international data streams. Retroactive data capture must be part of the standard operating procedures for collections at a scale that can be sustained. The open sharing of collections should be the rule, with the embargo of data for well justified and documented reasons only an occasional exception.

Research communities and other groups that are involved in the collection of specimens and data must stay abreast of best practices and standards governing the collecting of rich, augmented data sets through consultation with collections professionals at the initial phases of planning such collecting. For example: field notes, data spreadsheets, images, additional specimens outside of targeted species (“by-catch”), and possible ancillary environmental/ecological datasets should not only be gathered intentionally, but be compiled and stored in such a way that they are easily retrievable and link seamlessly to the core specimen metadata. Managing collections in light of these new data streams will require new best practices as well as tools, frameworks, and pipelines to accommodate these aspects of the extended specimen.

In a future scenario of greater collaboration and data exchange, collections professionals will divide their attention between addressing needs of their own collections and participating in activities of the greater collections community. Consequently, they will be well positioned to adjust administrative practices in their own collections to national norms. By developing the ability to place their own collection in a national context, collections professionals can help senior institutional leaders (e.g. provost, research administrators, trustees) understand the central roles that biodiversity collections play in supporting research. Realizing the value of collections will hopefully lead to institutional buy-in on funding campaigns to support education and research that is only possible because of an institution’s scientific collections.

Build or Strengthen Strategic Partnerships

The NIBA Strategic Plan called for the development of a web of partnerships among the stakeholders for digitized collections data to ensure the success of a digital collections network, and indeed such a web has developed in the past ten years (see Appendix B). To create the extended specimen network envisioned requires that we build on existing partnerships and create new ones.

Computer and data science research communities. Building the extended specimen network will require strong interdisciplinary development with the computer and data sciences community in order to build and maintain next generation collections infrastructure. Implementing some form of blockchain technology for specimen tracking and developing semi-automated approaches to data completion and standardization is another area of collaboration between the biodiversity collections community and computer science.

International Biodiversity Organizations. Surveys of the stakeholder community identified greater international collaboration as a key need for the next phase of collections data mobilization in the U.S. The Global Biodiversity Information Facility (GBIF) now in its 20th year of operation, recently surpassed one billion species occurrence records available for searching through their data portal. GBIF has served as a model and advisor for the implementation of the NIBA plan and the iDigBio data portal. Given their scope and vision, GBIF is arguably our most important international partner in the implementation of our new national agenda.

Among other national or regional data aggregators, Atlas of Living Australia (ALA), the most mature national biodiversity resource, provides an excellent model for developing user interfaces that meet the needs of the broader community. The Distributed System of Scientific Collections (DiSSCo) is a new European Union program that aims to "position European natural science collections at the centre of data-driven scientific excellence and innovation in environmental research, climate change, food security, health and the bioeconomy." Inspired in part by the ADBC program, DiSSCo is poised to explore new uses for collections data from which we can learn and in which we can collaborate.

At a local scale, we must pursue collaboration with data aggregators in Mexico (CONABIO) and Canada (Canadensys) to permit the seamless transfer of the data needed for continental-scale understanding of the breadth of biodiversity, its distribution and change over time. Common approaches among all three North American countries would help to fill gaps through new collections and digitization of existing ones. Opportunities for collaboration in the development of research tools and training programs would strengthen the capabilities in all three North American countries.

Aggregators of related data. Integrating historic occurrences and interactions of organisms with current observations is crucial for understanding environmental change. To do so will require new tools for data integration and analysis. Collaboration with programs such as the National Ecological Observatory Network (NEON), Critical Zone Observatories (CZO), and the Terrestrial Ecosystem Research Network (TERN) will help to establish standards and protocols for combining historical and current occurrence records to stimulate new investigations in outdoor laboratories, such as field stations and Long Term Ecological Research (LTER) centers.

Examples of other databases that enrich the context of biodiversity specimens include taxonomic resources such as the Catalog of Life checklist, published literature such as the Biodiversity Heritage Library, publishers of genetic information such as the Barcode of Life Data System (BOLD) and Genbank, and the Encyclopedia of Life Traitbank. As discussed in detail in Hobern, et al. (2018), these systems remain largely unconnected and do not yet function as a global infrastructure in the way needed to create extended biodiversity specimens that leverage the full value of existing data. Participation to the fullest extent possible in GBIF's proposed alliance for biodiversity knowledge will facilitate local work and help align our efforts to observe, measure and model life on earth with global efforts.

Professional Societies. Strengthening the relationship with taxonomically based societies (e.g., plants, vertebrates, insects, fungi), as well as collections and data oriented organizations (e.g., Natural Science Collections Alliance, Society for the Preservation of Natural History Collections and the Taxonomic Databases Working Group) will facilitate the greater use of collections data in research and training. In the case of taxon-oriented societies, greater involvement with the digitization process will inculcate a greater sense of responsibility for the protection and development of collections in their domain. Developing an efficient means of tracking specimens through their use in scientific publications will require partnership with publishers of primary scientific literature such as society journals as well as their editorial boards.

The American Institute of Biological Sciences (AIBS), a federation of scientific societies spanning the biological sciences, works to unify the community around common interests. Working with organizations like AIBS and iDigBio, the new agenda can further expand linkages to a wider diversity of scientific fields, journal editors, funders and policymakers.

Education and Broadening Participation. Working with national educators organizations as well groups such as the Society for Advancing Chicanos and Native Americans in Science (SACNAS) as well as Minority Serving Institutions (MSIs), Historically Black Colleges and Universities (HBCUs) and other higher learning institutions that primarily serve groups underrepresented in science, must be a central effort of future collections initiatives. Finding partners among agencies with programs to help people gain the skills necessary for self-sufficiency should be a priority for all collections institutions. Partnerships among robust and fledgling citizen science initiatives will promote projects that can be widely implemented with compatible data standards and protocols.

Other Partnerships. The development of partnerships beyond our primary user groups is key to sustaining both the national data store and the collections themselves. The stakeholder community identified an extensive list of the scholarly and business-oriented groups with which potential partnerships can be formed for knowledge sharing and financial support, summarized in Appendix C. It must be noted that many of these groups were also identified as potential partners in the NIBA plan. Although we have direct evidence of use of collections data by many of these groups, the extent of such use is unknown, and formal partnerships have not developed in most cases. An openness to collaboration with non-traditional partners in academia and industry could lead to wider application of pertinent technology. As the breadth of users and value and diversity of products generated from digitized collections data grows, so too will support for investments in biodiversity collections. As our extended specimen network becomes more representative of U.S. national biodiversity holdings, and the data are more standardized and thereby more easily searched and combined, we will be able to offer the most relevant and complete data sets to users, and owners of the data can be credited and compensated for their use. The first partnership that needs to develop in this context is with business development advisors who can help us with marketing as well as cost- and profit-sharing agreements.

V. Implementing and Sustaining the New Agenda

Building a master set of extended specimens that represent the wealth of biodiversity held in U.S. collections will require a monumental effort, comparable to building a new telescope for planetary exploration. However, the lack of physical infrastructure on the scale of a telescope makes the magnitude of the effort required to build and sustain such a resource harder to comprehend. A major challenge to the implementation of the new agenda and maintenance of the resulting information network will be to convey the unity of the resource and the scope of the effort required to build and maintain such a resource.

Initially building the infrastructure needed for the extended specimen network might be accomplished through established grant programs. Episodic funding for particular digitization or data augmentation projects would also be effective for filling knowledge gaps, improving data and for development of new educational initiatives. However, sustaining the extended specimen network requires a coordinating center funded over a much longer time horizon than currently supported by any existing grant program. Building upon the current model of the digitization hub (iDigBio) established in the ADBC program, a securely funded coordinating center could maintain the data network and partner with collections institutions and professional societies to support the stakeholder community to share techniques, resources, strategies for outreach and demonstrating the value of collections. We assert that the creation of a platform as ubiquitous and indispensable as the GenBank database, with a similar open-ended funding, is required for the extended specimen network to reach its full potential. Indeed, Genbank's data would be integrated within the larger framework of the proposed extended specimen database, thus further enhancing Genbank's utility as well. Although we recognize the investment required is large, the future of the planet's biodiversity may well rely on the knowledge we can gain from the mobilization and integration of these data.

Although securing the long-term funding base for the extended specimen network will take some time to develop, there are steps we can take now to set the process in motion, including the following:

- Develop a robust, comprehensive specimen identifier system in collaboration with other international data aggregators and providers. This will enable transparent and uncomplicated integration of biodiversity data with other data sources while facilitating the attribution of collections' role in discovery and policy and promoting transparency for broader issues involving multiple stakeholders (e.g., access and benefit-sharing).
- Create an authoritative, comprehensive, and self-updateable index of U.S. collections institutions (similar to Index Herbariorum for global herbaria) with structured metadata to describe their holdings as a first step toward expediting the discovery of undigitized collections and revealing these to the research community.
- Continue digitization of existing material focused on underrepresented taxa (e.g., those in entomology and paleontology) and including specimens held in small regional, personal, and individual researcher based collections. Additional efforts need to include improvement of previously digitized specimen data by imaging specimens, completing skeletal records, and augmenting data with georeferencing.
- Develop new protocols for the collection and accession of data-rich samples that provide greater context for understanding the biotic and abiotic interactions of organisms and create comprehensive datasets for research and education.
- Develop data access tools designed to maximize the educational potential of collections and collections-based data to inspire interest in the natural world and enable a diverse, digitally fluent workforce, and allow citizen scientists to contribute meaningfully to documenting Earth's biodiversity.
- Support enhanced training of professionals for interdisciplinary work in biodiversity, data science, and informatics. Emphasize skills needed to work collaboratively in interdisciplinary teams and other soft skills related to communication, creativity, and critical thinking as these skills are paramount to conducting transformative science, communicating research, and cultivating and leveraging relationships with new research and user communities.
- Continue the support of a coordinating center for the management of the nation's biodiversity data store and oversee the transition of the national collections database into an extended specimen network.

VI. The Extended Specimen Agenda and the National Science Foundation's 10 Big Ideas

If implemented, the new agenda for biodiversity collections proposed here will provide a scientific tool for biological sciences as well as a resource that enables progress toward cross-cutting and frontier challenges reflected by the National Science Foundation's 10 Big Ideas. For example:

Enhancing Science and Engineering through Diversity. The place-based capacity of collections specimens and associated data combined with the social and societal relevance of biodiversity science can serve a role in creating inclusive, culturally relevant, and socially conscious educational materials that engage a broad and diverse audience in biodiversity science. Collections institutions have great potential to engage a broad range of young people to create a scientifically literature work force and build the ranks of our scientific and engineering communities.

Understanding the Rules of Life. To understand the biological and environmental factors that influence the phenotypes of the wide range of organisms that humans depend on will require exploration of organisms across the tree of life. The ability to derive molecular data from preserved specimens creates an efficient source, even for rare or possibly extinct organisms. Specimens provide fundamental data to current and past efforts to construct the tree of life (e.g., OToL, AToL, AVaToL, GoLife), and, conversely, these efforts provide critical linked data for further extending these specimens.

The Future of work at the Human Technology frontier. The use of biodiversity collections in projects to both refine current techniques in machine learning and to document variation among organisms as exhibited by specimens holds great promise (e.g., Carranza-Rojas, et al., 2017, McAllister, et al, 2018). Use of these tools will lead to new avenues of data analysis, requiring new skills for researchers and data managers. Collections professionals will find themselves at this frontier and supporting their adoption of new technology and documenting the experience through training and best practices is a key aim of the agenda presented here, and the work done in this community may prove transferable to other communities.

Midscale infrastructure. The level of support required for the extended specimen network coordinating center falls within the gap between what NSF currently funds as either small and large infrastructure. The computer and human infrastructure required for data acquisition, deployment and training to support the extended specimen network represent a midscale infrastructure need, both in terms of funding level and the scientific research it will support.

Navigating the New Arctic. The Arctic biota sampled over the past two centuries and stored in our national collections provide a baseline for understanding the implications for life of the rapid rate of warming in this region that is critical to global economy and security. Digitized data from collections can provide historical context for the proposed network of observational platforms, and a reference for the identification, distribution, and behavior of species in the Arctic.

Harnessing data for 21st Century Science and Engineering. The digitization of natural history collections has already contributed to the "deluge of data" from the nation's scientific facilities, and as we broaden the knowledge bank of specimens with additional genotypic, phenotypic, and environmental data, the amount of data will increase exponentially. The extended specimen network will be a prime example of a "cohesive, national scale approach to research data infrastructure," and may inform efforts in other domains. In addition, specimen based data provides a unique opportunity to educate students in data science and train a data enabled workforce. Collections data alongside the archived specimens are an engaging and accessible data source that can provide students an opportunity to experience the entirety of the data pathway, while practicing verifiable and testable science.

Growing Convergent research at NSF. Enhancement of the roles that collections-derived data can play in understanding and protecting human health and in education are key objectives of the new agenda, to be achieved through the central efforts to integrate and share data more widely. The extended specimen network, combining diverse data sources linked to a physical museum specimen is a physical manifestation of convergent research and will amply demonstrate the power of this approach.

VII. References

AlShebli, B. K., T. Rahwan, W. L. Woon. 2018. The preeminence of ethnic diversity in scientific collaboration. Nature Communications. 9: 5163.

Ballard, H. L., Robinson, L. D., Young, A. N., Pauly, G. B., Higgins, L. M., Johnson, R. F., and Tweddle, J. C. 2016. Contributions to conservation outcomes by natural history museum-led citizen science: Examining evidence and next steps. Biological Conservation. https://doi.org/10.1016/j.biobon.2016.08.040.

Bebber, D. P. et al. 2010. Herbaria are a major frontier for species discovery. Proceedings of the National Academy of Sciences 107: 22169-22171.

Campbell, P. 2017. Archaeology and blockchain: a social science data revolution? https://www.theguardian.com/science/2017/oct/02/archaeology-and-blockchain-a-social-science-data-revolution Mon 2 Oct 2017 07.20 EDT

Carranza-Rojas, Jose, H. Goeau, P. Bonet, E. Mata-Montero, Joly, A. 2017. Going deeper in the automated identification of herbarium specimens. BMC Evolutionary Biology 17: 181.

Contreras, D.L. 2018. A workflow and protocol describing the field to digitization process for new project-based fossil leaf collections. Appl Plant Sci 6:e1025.

Cook, J., et al. 2014. Natural history collections as emerging resources for innovative education. *BioScience* 64: 725-234.

Halamka, J. D., A. Lippman, Ekblaw, A. 2017. The Potential for Blockchain to Transform Electronic Health Records. Harvard Business Review. https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records (accessed 28 November 2018).

Hartop, E. A., Brown, B. V., and Disney, R. H. L. 2016. Flies from L.A., The Sequel: A further twelve new species of Megaselia (Diptera: Phoridae) from the BioSCAN Project in Los Angeles (California, USA). Biodiversity Data Journal 4.e7756

Hilsberg, T., Robinson, A., Robinson, M., Haynes, D. 2018. Biocoin.Life Foundation. White Paper V0.85. https://static1.squarespace.com/static/5a338858d7bdcea6d17ccc0f/t/5a7f8d45652deaf340ccdd98/1518308678928/White+Paper+-+Developing+a+Global+Biocoin.life+20180211.pdf

Hobern, D., et al. 2018, GBIC2, Towards a coordination mechanism for Biodiversity Informatics. In press.

Kemp, C. 2015. The endangered dead. Nature. 158:292-294.

Kissling, W.D., Walls, R., Bowser, A., Jones, M.O., Kattge, J., Agosti, D., Amengual, J., Basset, A., Van Bodegom, P.M., Cornelissen, J.H. and Denny, E.G., 2018. Towards global data products of Essential Biodiversity Variables on species traits. *Nature ecology & evolution*, p.1.

Marshall, C.R. 2018. Digitizing the vast 'dark data' in museum fossil collections https://phys.org/news/2018-09-digitizing-vast-dark-museum-fossil.html#jCp)

Marshall C.R,, S. Finnegan, E.C .Clites, P.A. Holroyd, N. Bonuso, C. Cortez, E. Davis, G.P. Dietl, P.S. Druckenmiller, R.C. Eng, et al. 2018 Quantifying the dark data in museum fossil collections as palaeontology undergoes a second digital revolution. Biol Lett 14:2–5. https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2018.0431

McAllister, C.A, M. R. McKain, M. Li , B. Bookout & E. A., Kellogg. 2018 Specimen- based analysis of morphology and the environment in ecologically dominant grasses: the power of the herbarium. Phil. Trans. R. Soc. B 374: 20170403.

Monfils, A., Powers, K., et al. 2017. Natural History Collections: Teaching about Biodiversity Across Time, Space , and Digital Platforms. Southeastern Naturalist 16: 47-57. https://doi.org/10.1656/058.016.0sp1008.

Powers, K.E., et al., 2014. Revolutionizing the Use of Natural History Collections in Education. Science Education Review, 13 (2): 24-33 2014

Schindel, D.E. & J.A. Cook. 2018. The next generation of natural history collections. PLoS Biol 16(7): e2006125.

Seltmann K., et al. 2018. *Georeferencing for Research Use* (GRU): An integrated geospatial training paradigm for biocollections researchers and data providers. Research Ideas and Outcomes 4: e32449. https://doi.org/10.3897/rio.4.e32449

Singer R.A., Love K.J., Page L.M. 2018. A survey of digitized data from U.S. fish collections in the iDigBio data aggregator. PLoS ONE 13(12): e0207636. https://doi.org/10.1371/journal.pone.0207636.

Spear, D. M., Pauly, G. B., and Kaiser, K. 2017. CItizen Science as a Tool for Augmenting Museum Collection Data from Urban Areas. Frontiers in Ecology and Evolution 5:86.

Vendetti, J.E., Burnett, E., Carlton, L., Curran, A.T., Lee, C., Matsumoto, R., Mc Donnell, R., Reich, I., and O. Willadsen. 2018. The introduced terrestrial slugs Ambigolimax nyctelius (Bourguignat, 1861) and Ambigolimax valentianus (Férussac, 1821) (Gastropoda: Limacidae) in California, with a discussion of taxonomy, systematics, and discovery by citizen science. Journal of Natural History. https://doi.org/10.1080/00222933.2018.1536230.

Webster, M.S. 2017. The Extended Specimen: Emerging Frontiers in Collections-based Ornithological Research. Boca Raton, FL: CRC Press/Taylor & Francis Group

VIII. Authorship of this document

Thiers, Barbara, Anna Monfils, Jennifer Zaspel, Andrew Bentley, Elizabeth Ellwood, Katherine Levan, John Bates, David Jennings, Dorilynne Contreras, Laura Lagomarsino, Linda Ford, Robert Guralnick, Robert Gropp, Marcy Revelez, Neil Cobb, James Lendemer, Katja Seltmann and Mary Catherine Aime.

IX. Appendices (Separate Documents)

Appendix A: Workshop Participants

Appendix B: NIBA Digitization Progress

Appendix C: Stakeholder Outreach Summaries

Appendix D: Selected References on Recent Uses of Biodiversity Collections