Manuscript for submission to *BioScience – Biologist's Toolbox*

Designing Biorepositories to Monitor Ecological and Evolutionary Responses to Change

KELSEY M. YULE, EDWARD E. GILBERT, AZHAR PALIT HUSAIN, M. ANDREW JOHNSTON, LAURA ROCHA PRADO, LAURA STEGER, AND NICO M. FRANZ

The authors are affiliated with the School of Life Sciences at Arizona State University, in Tempe, Arizona.

Author names, titles, and e-mail addresses:
Dr. Kelsey M. Yule – kelsey.yule@asu.edu (Corresponding Author) 0000-0002-1447-849X
Edward E. Gilbert – egbot@asu.edu 0000-0002-5006-8092
Azhar Palit Husain – azhar.husain@asu.edu 0000-0002-0639-4093
Dr. M. Andrew Johnston – ajohnston@asu.edu 0000-0002-0166-6985
Dr. Laura Rocha Prado – lauraprado@asu.edu 0000-0003-1237-2824
Laura Steger – lsteger@asu.edu 0000-0002-0878-4132
Dr. Nico M. Franz – nico.franz@asu.edu 0000-0001-7089-7018

Author affiliation and postal address:
Biodiversity Knowledge Integration Center and Biocollections
School of Life Sciences
Arizona State University
734 W. Alameda Drive
Tempe, AZ 85282 U.S.A.

Abstract: *We discuss the prospects of developing biorepositories as a novel kind of biocollections infrastructure and service. Biorepositories resemble other natural history collections but their structured material sample composition is optimized to monitor and forecast the dimensions of ecological and evolutionary responses to change. Samples are collected to study shifts in genetic composition, species populations and traits, community composition, and ecosystem function and structure. This concept is exemplified in the design and operations of the National Ecological Observatory Network (NEON) Biorepository. To maximize impact, biorepositories have to establish practices for sample collection, preservation, diversified high-volume use, and innovative data science that both leverage and diverge from more* *conventional collections management strategies. In light of their constrained purpose and transcendent scale among material sample collections, biorepositories face special challenges to generate a meaningful, inclusive sense of community ownership. Properly designed and delivered, they can catalyze broader support for biocollections and open science.*

Keywords: *biodiversity informatics, ecological monitoring and forecasting, global change, natural history collections, open science*

Natural history collections (see Box 1 for definitions of terms) are increasingly being recognized as underutilized resources for addressing ecological and evolutionary research challenges with broad societal impact. They provide access to preserved specimens and material samples (Van Allen 2018) covering spatial and particularly temporal scales that are otherwise inaccessible to individual research teams (Pyke and Ehrlich 2010, Johnson et al. 2011). Collections-based research can allow for direct observation of how populations and communities have responded to decades, if not centuries, of anthropogenic global change (Kharouba et al. 2018, Meineke et al. 2018, Heberling et al. 2019). Heberling and Isaac (2017) emphasize that specimens in natural history collections are "exaptations": originating in the context of biodiversity discovery and sustained exploration, specimens are frequently being co-opted for new uses as both needs and technological capabilities expand. These novel, critical roles include revealing genomic signatures of selection (Holmes et al. 2016), rapid morphological change (Weeks et al. 2019), phenological shifts (Jones and Daehler 2018), changes in nutrient availability (McLauchlan et al. 2010), and pathogen spread (Yoshida et al. 2014).

Co-optability of natural history collection specimens for these new uses is not without limits, however. The suitability of specimens to deliver strong, clear data signals relative to specific research needs is connected to the motivations and actions that generate them. For instance, collections growing out of a primarily systematic, discovery-focused legacy tend to project a static view of biodiversity (Ward 2012, Meyer et al. 2016, Daru et al. 2018). Voucher-based floristic or faunistic inventories are often designed to discover new taxa and understand the presence of known species in focal areas. Unless regularly updated, they are less well suited for understanding community turnover (but see Fernández-Triana et al. 2011 for an example that leverages a resampled site). Within an exploration-centric sampling paradigm, conspecific individuals are rarely preserved at sufficient temporal density to understand population-level trends in either abundance or phenotypic and genotypic variation (but see Weeks et al. 2019 for an example using well sampled species). Additionally, such collecting efforts are unlikely to provide sufficient spatial density to distinguish between the lack of adequate prior sampling and actual shifts in range (Frey 2009). Some constraints can be statistically mitigated; e.g., through the generation of pseudo-absence data for species distribution models (Ponder et al. 2001, Lütolf et al. 2006). On the other hand, statistical methods cannot overcome an absence of well-preserved specimens – whether extended (Webster 2017) or holistic (Cook et al. 2017) – for the purpose of recording species associations such as host-parasite relationships.

Biobanks or biorepositories (see Box 1 for definitions) are concepts most regularly used in the biomedical domain (Baker 2012). These collections preserve and maintain associated data for biological samples ranging from DNA and tissues to whole organisms, with an explicit mandate to redistribute these to researchers (Fullerton et al. 2010).

We examine a concept and implementation of a biorepository that focuses on non-medical (non-human), organismal and environmental samples. Although the taxonomic composition and the means of sample processing and preservation broadly overlap with traditions of natural history collections or biobanks, this concept is nevertheless novel and relevant because of the highly structured, consistent means with which the samples are generated. The primary intent and optimization strategy for sampling is not to discover "standing biodiversity" but to discover and characterize the dimensions of responses to ecological and evolutionary *change.* This focus would seem deliberately narrow, or even out of scope, in the context of natural history exploration (Schilthuizen et al. 2015). Optimizing the sampling of biodiversity so as to best monitor the dimensions of change immediately forces us to consider a range of feasibility and impact trade-offs, because one cannot resample everything, everywhere, indefinitely and consistently. Moreover, the mandate to detect change has unique implications for downstream process design and management. A well-run biorepository in the present sense must be *proactive* in redistributing in a timely manner as many samples to as many diverse research streams as possible. These characteristics make change detection-optimized biorepositories a new component of the global set of biocollections (Box 1).

It is our purpose to outline the opportunities and challenges that biorepositories may bring about, inherently and particularly in relation to a global evolving landscape of biocollections and biodata science. Our familiarity with the NEON Biorepository, for the National Ecological Observatory Network (Keller et al. 2008, Thorpe et al. 2016, SanClements et al. 2020), will serve to illustrate the realities and potential of change-focused biorepositories in this greater context.

Dimensions of monitoring ecological and evolutionary change

Biorepositories tasked with monitoring change operate under a productive, necessary tension that affects global-change science generally. On one hand, there are strong incentives to globalize and centralize the content and production of a consistent, constrained set of indicators (e.g., Pereira et al. 2013). On the other hand, there are sensible warnings against a one-size-fits-all approach to linking indicator signals and decision making; in favor of custom service, co-produced models are more effective in closing the feedback loop between data and decisions but are also more challenging to scale (Cash et al. 2006, Zulkafli et al. 2017, Lemos et al. 2018). Biorepositories can take either approach, but we will pragmatically review how they can contribute to the measurement of six high-level classes of indicators promoted by the Group on Earth Observations Biodiversity Observation Network (GEO BON, Scholes et al. 2012; see also https://geobon.org/ebvs/what-are-ebvs/). In doing so, we will show that biorepositories can be designed to allow for direct interrogation of population, community, and ecosystem level processes, thus becoming a central resource for monitoring and forecasting of ecological and evolutionary change.

Genetic composition. Biorepositories hold material samples enabling future researchers to observe how species' genetic compositions respond to global change. Variables to monitor include: co-ancestry, i.e., pairwise relatedness and inbreeding coefficients between individuals; allelic diversity, i.e., allelic richness across loci and populations; and population genetic structure, i.e., differentiation among populations. They capture the ability of species to adapt to novel challenges and the extent to which populations are genetically distinct, cryptically diverse, or reliant on genetic connectivity (Bálint et al. 2011; Caballero and García-Dorado 2013). Accordingly, biorepositories need to preserve enough samples from a recognized species to adequately reflect within-species changes in genetic composition and variation. This entails structured, repeated sampling of the same species at the same geographic sites. In many cases, small numbers of conspecific individuals per sampling locality and event suffice for population genomic studies leveraging high-throughput sequencing technologies (Willing et al. 2012). However, to minimize geographic bias towards particular populations, sampling efforts that are distributed proportionally across putative populations will be most useful for biorepository preservation.

Species populations. By preserving many conspecific individuals, biorepositories also facilitate population dynamical and demographic trend analyses. Regardless of sampling strategy, population abundance and structure variables – i.e., distributions of age and size classes – are often difficult to address; yet they are critical for assessing extinction risks, pest outbreaks, or invasive species spread. Most important for monitoring purposes is a well-specified and consistent sampling effort. By supporting the systematic resampling of locations across time, biorepositories will allow researchers to estimate abundance trends without making assumptions about sampling effort or the relationships between community-wide and relative abundances (Jeppsson et al. 2010, Bartomeus et al. 2013). Given sufficient temporal sampling density, changes in the composition of population age, sex, or size class structure can also be determined (de Paula and Melo 2015). As they often document the presence of individuals from low density, range-edge populations, natural history collections can be better suited to studies of species distribution; however, species distribution models that incorporate interpolated abundances, as could be determined from biorepository sampling efforts, perform better than those using presence-only data (Gomes et al. 2018). Furthermore, biorepository samples may improve forecasts of climate-change induced range shifts by providing the records needed to study adaptability, biotic factors, and population dynamical responses to co-located abiotic fluctuations (Lewis et al. 2017).

Species traits. Archival of many conspecific individuals enables studies of phenotypic changes through time. Biorepositories are particularly powerful for the study of trait variables concerning intraspecific variation in phenology and morphology (Kissling et al. 2018). Trends in phenology are used to understand species' sensitivities to climate change and the potential for climate-induced mismatches among interacting species (Thackeray et al. 2016). Although there are powerful methods for tracking phenology through collection records (Jones and Daehler 2018), biorepository samples are less prone to seasonal sampling biases that can affect the accuracy of these methods. Phenotypic analyses using specimens to document body size reductions can indicate selection pressures on species in response to overharvesting and warming (Taylor et al. 2006, Weeks et al. 2019). Structured biorepository sampling can facilitate those trait-based studies by preserving conspecific variants relative to their frequencies. In contrast, natural history collections often preferentially sample individuals of markedly different phenotypes, not in close relation to their relative frequencies.

Community composition. To best provide insights into community-level changes, biorepositories must preserve samples that are representative not just of populations but of ecological communities; i.e., samples that capture the identities and relative frequencies of species and their interactions. Sampling and sample curation methods should therefore reflect the richness and diversity of communities to support monitoring taxonomic diversity as well as species interactions. Taxonomic diversity is used to classify habitats, ecosystem stability, and to identify signals of community turnover (Laurance et al. 2006). While natural history collections are often more likely to contain rare species (Garcillan and Ezcurra 2011), biorepositories can overcome their inherent limitations for estimating evenness and diversity, through consistent sampling that is unbiased relative to species' densities. Monitoring species interactions is relevant to understanding the persistence, ecosystem services, and resistance of communities to invasions (Tylianakis et al. 2010). Biorepositories can preserve specimens that show damage indicative of species interactions – such as plants damaged by herbivores – in proportion to their occurrence in the population, even though these may traditionally be considered low-quality vouchers (Cook et al. 2017, Meineke et al. 2019). Similarly, without even retaining whole organisms, biorepositories can record changing species interactions; for instance, changes in dietary relationships revealed through stable isotopes in hair and bone samples, as well shifts in pathogens and parasites in blood samples and microbiomes in mammalian feces (McLean et al. 2016).

Ecosystem function. Biorepositories are particularly well suited among biocollections for documenting ecosystem function; i.e., the collective biological, chemical and physical processes that determine the flows of matter and energy through an ecosystem. Biorepositories can focus on documenting biotic and chemical characteristics of the environment at sites for which measurements of other ecosystem factors – including nutrient inputs, primary productivity and carbon fluxes – are being produced synchronously. The coordinated sampling enables studies of nutrient loss, retention, and biochemical cycling more generally, which are critical to the productivity and services provided by ecosystems and the levels of biodiversity they support (Penuelas et al. 2020). Extensive simultaneous preservation of environmental (e.g., soil, water, atmospheric deposition), microbial, primary producer (e.g., litter, belowground biomass, canopy foliage), and higher trophic level (e.g., mammal hair) samples in biorepositories will facilitate analyses of matter and energy exchanges through an ecosystem's a-/biotic components. This includes understanding carbon, nitrogen, and phosphorous cycling and heavy metal accumulation (McLauchlan et al. 2010, Ziska et al. 2016, Rudin et al. 2017).

Ecosystem structure. Biorepositories can support characterizations of an ecosystem's physical and trophic structure, size, inter-system flows, and of the spatiotemporal variability in a-/biotic pools of matter and energy. They supply samples used towards monitoring disturbances, fragmentation, overharvesting, and changing climates; hence allowing for assessments of functional diversity and/or functional dominance, which can predict the productivity and stability of an ecosystem and its component communities of interacting species (Fontaine et al. 2006, Pérez-Ramos et al. 2017, Zhang et al. 2017). Sometimes these variables can be estimated simply by measuring phenotypic traits readily obtained from specimens (e.g., specific leaf area; Loranger and Shipley 2010). Such analyses require that samples are taken from individual organisms that in aggregate represent the functional composition of the community at a point in space and time. In other cases, the functional composition of communities (e.g., soil and aquatic microbial communities) can be obtained from biorepository bulk samples using metagenomic or metatranscriptomic methods (Escalas et al. 2019).

Designing impactful biorepositories

We now shift from an overview of novel and enhanced capabilities of biorepositories to monitor change – relative and complementary to natural history collections – to strategic considerations for designing and managing them to maximize impact.

Sample collection. As reviewed above, biorepository samples provide lasting material evidence towards monitoring broad dimensions of ecological and evolutionary change. They must therefore accumulate and in turn redistribute samples which are representative of intra- and interspecific variation and produced from strictly controlled and repetitive collection efforts at the same set of localities. As long as there is sufficient, higher-volume sampling effort, community- and population-level trends are discoverable by following the same sampling approach. Large collections of common species are likely among the most valuable samples for the purpose of monitoring change, while also being unlikely to duplicate standing efforts of natural history collections.

Sample collection will ideally occur both within and between years, in order to capture shifts in phenology and avoid confounding other biodiversity measures with either species-level differences in phenology or population fluctuations. The sampling will be most informative when conducted across large geographic and broad taxonomic dimensions. However, this requires managing feasibility trade-offs: we recommend prioritizing high-depth (i.e. large numbers of thematically similar samples) and frequent resampling of a limited set of localities and with a narrow taxonomic focus. Pairing the organismal sampling effort with detailed, fine-scale environmental data production is particularly critical for direct comparisons of abiotic and biotic fluctuations. Repetitive, consistent sampling regimes are most conducive to answering conceptual questions about responses to global change. If such narrowly defined efforts are deployed by many biorepositories, then in aggregation they enable the discovery of change across regional and taxonomic dimensions.

Sample preservation. Population- and community-wide collecting strategies require far more samples and specimens to be preserved than might be feasible or desirable in natural history collections. Faced with constraints on support for personnel managing and processing samples (including taxonomic experts carrying out identifications), time, and collections space and infrastructure, biorepositories can relax conventional practices under which every individual organism would be separately prepared and further curated. As long as samples are clearly linked to the collecting event data, multiple specimens may be provisionally contained in various forms of "bulk" storage. Additional preparation and analysis of individual specimens from bulk samples can then follow when dictated by external research interests. Retaining bulk non-target specimens ("bycatch") from structured sampling efforts can be particularly impactful – leading to species discovery, updated distribution data, regional inventories, macroecological insights, and other products – while limiting additional fieldwork and supply costs (Buchholz et al. 2011, Spears and Ramirez 2015). For some bulk sample types, conventional processing and identification of individual specimens may be bypassed through metagenomic and environmental barcoding methods that provide diversity estimates even for organisms that were not actually collected (Linard et al. 2016, Lyngaard et al. 2019).

Where possible, biorepositories should additionally preserve "extended specimens" (Webster 2017) and environmental samples, as these will engage a wider variety of researchers whose focal research may typically not include natural history collection specimens. Samples of extracted DNA, tissue, blood, feces, hair, toenails, bycatch, soil (etc.), will provide the raw materials for answering future researchers' genomic, population, and ecosystem-level questions. Many of the latter samples require cryo or ultralow storage conditions – via extensive liquid nitrogen and/or mechanical freezer infrastructures – that are not yet well developed in many natural history collections facilities.

Sample use. We further discuss this central topic in the section on challenges and opportunities. Generally speaking, and relative to the majority of natural history collections, established biorepositories will operate under a stronger linkage between justifying their existence and the volume, frequency, diversity, and timing of sample use. Biorepositories will commensurately succeed by redistributing samples to researchers for further analysis of change signals – at any moment and in any dimension where significant change may actually occur. To avoid delaying responses to change signals for decades after they would have been possible and relevant (cf. Fontaine et al. 2012), frequent, abundant, and diverse forms of biorepository use must be actively promoted and realized. These justified uses will on balance involve a higher degree of sample alteration, and partial or entire sample consumption (to full destruction). Transparent policies that facilitate variously destructive types of usage will foster engagement by a broad set of ecologists and evolutionary biologists whose research methods might otherwise preclude access to more individually valuable specimens in natural history collections.

However, because biorepositories also document longer-term change, the benefits of immediate destructive sample use must recurrently be weighed against future signal discovery potential (Ayres 2019). An evolving relationship between short-term realization and longer-term reduction of research impact, in effect favors sample use policies that are more flexible, and more dynamic than a generic approach. Early and wide-ranging usage is critical as it establishes data signal baselines and encourages further investment in biorepositories and novel research directions. Conversely, methodological advancements and longer-term sample/data series – the deeper, the more powerful – mean that the data-production potential for any given sample will invariably increase over time. Recent analyses of natural history specimens with powerful genomic and machine learning methods (Rowe et al. 2011, Carranza-Rojas et al. 2017) support such a bet hedging strategy for use.

Most importantly, whenever samples are altered in any way, the details of that usage – e.g., potential sources of contamination – must be recorded and the resulting, value-added data must be linked to easily accessible online sample records published by data portals (as described below). To best navigate the counter-acting needs for heavy sample usage and longer-term sample retention for research, biorepositories must set innovative policies that define what data should be preserved in the event of destructive use. As an example, approval for destruction of an insect specimen could entail publishing a set of high-resolution images, DNA extract, and all of the project's target data within the respective biorepository data portal.

Data science. No aspect of developing impactful biorepositories requires more innovation than the domain of data science, broadly defined. The needs reach well beyond providing the baseline of FAIR Principles (Wilkinson et al. 2016). The more onerous goals aim at shifting incentives for research teams to encourage use of biorepository samples. Few researchers will proactively visit biorepositories in person to see what samples might be there that could potentially be fit for advancing their next project, perhaps partially due to a general lack of information on what such collections entail. Forward-looking data science, or biodiversity informatics more narrowly, will undoubtedly become the prevalent means of reaching and recruiting new biorepository sample/data users and contributors. This is the new frontier.

Suitable strategies will adhere to Open Data standards, by providing data that is not only accessible and portable, but has been assessed for quality and provenance, is accompanied by metadata, and enforces restricted vocabularies to provide semantic interoperability (W3C 2017). Biorepositories will create open, *easy-to-find-and-use* – not merely accessible – data portals, which are dynamic web applications that allow for publishing, editing, managing sample data, including linking to all associated, value-added research data (Lendemeer et al. 2020); such as phenotypic traits (Hedrick et al. 2020), species interactions (Wilson 2020), co-located species checklists (Johnston et al. 2018), and genetic information (Gibson et al. 2012). Sample data will adhere to evolving community interoperability standards (e.g., Darwin Core; Wieczorek et al. 2012, see Box 1), thereby facilitating publication and aggregation with a maximally broad spectrum of biodata communities and initiatives. This further entails the design and implementation of robust web-based services (such as Application Programming Interfaces) to expose and support data annotation, e.g. value-added data edits, and maintenance of data versioning and tracking provenance through the participation of an expanding network of collaborators. The data design, portal management interfaces, and programmatic exposure of data should all support the augmentation of post-collection, value-added data definitions (e.g. defining species associations, phenology, or phenological metrics based on collections material). Ideally, the design would support and even encourage the scientific user community to submit and define additional annotations and expanded specimen data definitions that increase the value of the basic specimen records. Interoperability should support the ability to define external value-added datasets that are programmatically mapped directly back to the original source records.

The NEON Biorepository as a test case

We describe the NEON Biorepository as a full-fledged example of the potential and constraints related to designing and operating biorepositories more generally.

Scope and implementation. The NEON Biorepository (NEON 2019) was established in 2018 at Arizona State University (henceforth: ASU), co-located with the ASU Natural History Collections, they now jointly comprise the ASU Biocollections. At present, NEON operates 81 field sites – 47 terrestrial and 34 aquatic – across 20 ecoclimatic domains and associated facilities distributed throughout North American subcontinent and including Alaska, Hawaii, and Puerto Rico (Keller et al. 2008, Kampe et al. 2011, Kao et al. 2012, Thorpe et al, 2016, Sanclements et al. 2020). Sampling and data protocols are highly structured and consistent across sites, and currently 180 data products are openly available through the main NEON data portal (https://data.neonscience.org/). They include high-resolution measurements of abiotic variables, remote sensing data, and organismal observational (non-material, see Box 1) data spanning multiple trophic levels.

Many of the NEON data products are related to the nearly 70 sample types that are housed at the NEON Biorepository (Fig. 1; Supplemental Material: Table 1). These material samples are taken at the field sites, then variously processed at NEON domain facilities and at external service labs, and finally shipped to and preserved at the NEON Biorepository (Fig. 2) . Although selective "legacy" sampling for NEON dates back as far as 2012, full operations across all field sites began in 2019 and are expected to continue for 30 years. From 2019 onwards, over 100,000 samples will be ingested annually (Fig. 1; Supplemental Material: Table 1), for a projected grand total of over 3 million samples. We estimate that the number of macro-organismal *specimens* contained in these samples is many times higher, due to standing practice to preserve bulk samples, including invertebrate "bycatch" from pitfall traps (Hoekman et al. 2017).

Figure 1 provides a high-level breakdown of the annual totals according to taxonomic category, sample type, and storage condition. The taxonomic composition is both broad and focused; for instance, ground beetles, mosquitoes, and small mammals are particularly well represented and identified. Many other groups are initially preserved as bulk, community-level samples with unrealized data potential. A large proportion of organismal samples are not stored as whole organisms but are instead DNA, blood, tissue, or feces. The NEON Biorepository also accessions environmental samples such as dry soils, plant belowground biomass, and particulate mass filters. Storage conditions for each sample class are chosen to optimize future data potential; hence the majority of samples (~60%) are maintained at cryogenic or ultralow temperatures.

Sample access and data publication. The NEON Biorepository’s public data portal (https://biorepo.neonscience.org/) serves as the primary access point for material sample data particularly in the context of research project development, sample loan facilitation, and dynamic data publication. This portal complements the main NEON data portal in at least two functional dimensions; both of which are related to the choice of the Symbiota software (Gries et al. 2014) as the portal's content management system. Symbiota is an open-source specimen management system that can: 1) be mobilized on the web, fulfilling interoperability requirements, 2) support the integration of both live-managed and remote snapshot collections managed by external collections, 3) and support a versioned annotation system that provides the potential for remote community members to participate. Therefore, the choice of Symbiota builds a cultural data science bridge to global biocollections and biodiversity data scientists. Symbiota is by design optimized for networking data for regionally, taxonomically, or otherwise socially constrained *communities of practice* (Wenger 2000). Symbiota serves natural history collections networks that self-assemble via thematically overlapping portal communities. Adherence to the Darwin Core standard and the issuing of Globally Unique Identifiers (GUIDs; Guralnick et al. 2015) for NEON material samples, are further promoters of data integration and publication with these communities (Edwards 2004). Additionally, Symbiota is designed to accredit and incentivize various forms of individual or group contributions aimed at enriching the quality and diversity of data related to NEON material samples. The NEON Biorepository data portal therefore has the potential to enable a continuous, collaborative and cyclical process of data annotation and publication relative to sample-based research projects (Fig. 2, purple arrows connecting portal and researchers). The ability to incorporate new data and annotations will require further refinement of web services and data infrastructure to harmonize data additions across a more inclusive, de-centralized network that must entail the main NEON data portal and other, third-party data publishers. We aim for researchers to be a central hub in the data flow hub.

Generating community ownership

The paramount opportunity for developing impactful biorepositories is to create a meaningful, inclusive sense of ownership. Because of the superficial resemblance with other biocollections, the degree of novelty or foreignness of resources such as the NEON Biorepository may be underestimated.

The NEON Biorepository is a biocollection with a distinct, federally supported mandate. It was designed as a community resource, transcending many conventional constraints for generating and redistributing samples to monitor responses to change. Reaching this state required sustained, pivotal contributions by many members of the ecological and evolutionary research community (cf. Dalton 1999, Senkowsky 2003, Keller et al. 2008). Now that the design has materialized and is operated by (in effect) third-party contractors, a shared, community-level history of biorepository scoping is in itself not enough to create a strong sense of biorepository sample and data ownership for individual contemporary researchers.Therein lies the challenge. Because of the necessarily tenuous relationship between the historical process of community-level conception and the resulting manifestation of the NEON Biorepository, reconstituting a meaningful sense of biorepository ownership now becomes our central, multi-dimensional task.

In addition to concerns about ownership, researchers may be hesitant to use Biorepository resources because they are: (1) unfamiliar with the processes by which biocollections samples and the data arising from them can be accessed and disseminated; (2) concerned about an inability to personally and directly oversee the sampling and data publishing protocols; and (3) unsure of how to weigh use of a new untested resource against their own established research program. These are generally valid points about intellectual ownership and trust (Franz and Sterner 2018). With appropriate filtering, such evolving concerns must function as guard rails leading to better biorepository operations. There is a fine line between 'for everybody' and 'not mine in particular'. The need to develop biorepository ownership is an upfront cost that comes with the narrower, divergent purpose relative to other biocollections, executed at a scale that exceeds most monitoring projects yielding material samples (such as Long-Term Ecological Research programs and sites; Hobbie et al. 2003). The most effective way to mitigate these costs is to regularly receive feedback from and remain engaged with the communities of practice that biorepositories serve. The NEON Biorepository solicits and welcomes this participation so that it can evolve to continue to be as useful and relevant as possible as needs change.

Strategies for engagement

We offer a number of recommendations for strategies that are most effective in lowering thresholds for biorepository engagement. The first is to allocate outstanding research support services towards early adopters. Success tends to breed success. Researchers who proactively 'dare' to engage with biorepositories despite the aforementioned sense of novelty and risk, must be involved in highly rewarding and empowering experiences that lead to impactful research outcomes. The success stories, and their pioneering narrators, will provide strong 'insider' cases in favor of broadening ownership of biorepository samples and data. Every instance of sample discovery, loan request fulfillment, and eventual derivative sample and data product return, is an opportunity to grow the community and create sustainable data cycles where biorepositories function as source and sinks for compelling research activities and outcomes.

The second is innovation in data science. Biorepositories are tasked with helping societies understand where, how, and why multi-dimensional ecosystems are (most) susceptible, or (most) resilient, in their ecological and evolutionary responses to change. Biorepositories can lower the thresholds that trigger research proposals and sample loans by providing novel services, such as preliminary assessments of fitness-for-purpose of samples and data. Further, because biorepositories are paradigmatic providers of "extended" samples and specimens (Lendemeer et al. 2020), data portals must be socially designed to optimize a sense of shared ownership of sample data quality and derivative data products. This means, minimally, that individual, novel contributions of value-added portal data must be feasible by design, linkable to primary authors and supporting sources of evidence, and permanently retrievable. Moreover, biorepository sample-related data must be reconfigured and republished – dynamically, through robust web infrastructures, and with provenance – for a maximally broad spectrum of external data aggregators and prospective research communities (e.g., Hobbie et al. 2003, Karan et al.

2016, Waide et al. 2017, Dietze et al. 2018, Hobern et al. 2019, Jetz et al. 2019, Larson et al. 2020).

As a third component of promoting a new sense of ownership, we stress education; i.e. the need to develop broad-based knowledge, curiosity and competency about biorepository samples and data. It is critical to promote this awareness not in isolation or through unproductive competition ("zero sum games"). Biorepositories likely have few functions that are not at least partially fulfilled by other existing resources. Conversely, integrating biorepository data will enrich many existing analysis pipelines. Our best prospects lie in fine-tuning selective integration. Because biorepository samples and data are new additions to an existing set of tools, education should explore reciprocal and complementary strategies to broaden and sharpen data sources and signals relative to research interests. With regards to the physical and information preservation and management culture, the NEON Biorepository is most closely aligned with natural history research collections. Yet in terms of research culture, it needs to primarily serve analytical, hypothesis- and model-driven ecological and evolutionary macrosystems scientists. Education can be creative and inclusive in leveraging the respective strengths of these cultures (Pyke and Ehrlich 2010, Johnson et al. 2011, Kharouba et al. 2018, Meineke et al. 2018, Heberling et al. 2019).

Lastly, we hope that biorepositories will leverage broader trends towards open, purpose-based science, where diverse but necessary roles and responsibilities for providing system- and community-level research infrastructures, data products, and related data science services increasingly achieve parity with traditionally recognized, career-making contributions (e.g., Allen and Mehler 2019, Leonelli 2019, Strasser 2019). Preceding large-scale monitoring efforts with overlapping goals have become indispensable data sources (e.g., Hobbie et al. 2003, Karan et al. 2016, Waide et al. 2017). Biorepositories can play a catalyzing and mutualistic role towards increased support for biocollections infrastructures and services (Suarez and Tsutsui 2004, Ward et al. 2015, Mclean et al. 2016).

What is the special value of constraining the spatial, temporal, and sampling scopes of biorepositories so deliberately upfront? How strong is the linkage between the predetermined sampling regime and the ability to utilize samples and data for in- or out-of-scope research questions? How can signals from biorepositories be expanded in spatial, temporal, and organismal coverage through addition through data from other biocollections? How should emergent signals from the latter inform the development of new or modification of operational biorepositories? What other components of research infrastructures and cultures are missing to mitigate shortcomings inherent in any distinct, realized biocollection type, relative to shared focal monitoring needs? We believe these are the kinds of new questions that developing biorepositories raise, and which can elevate an open, inclusive, community-enabled and -enabling approach to the science of monitoring responses to change generally.

In the Supplemental Material, we provide specific information to develop engagement with the NEON Biorepository. Upcoming, or well established generations of researchers now have an opportunity to "vote with their feet", in favor of open science principles and trade-offs that are also increasingly reflected in how biocollections are designed and operated (cf. Damalas et al. 2018).

Acknowledgements

The authors acknowledge the National Science Foundation for supporting operations of the National Ecological Observatory Network (NEON), managed under a cooperative support agreement to Battelle Memorial Institute (NSF EF-1029808), which in turn subcontracts with Arizona State University to operate the NEON Biorepository. We thank NEON's many collaborators and site-hosts for their commitment to realizing the Observatory. Christine Laney, Katherine LeVan, Paula Mabee, and Katherine Thibault provided invaluable input into the design and operations of the NEON Biorepository, as well as comments that improved this manuscript. We also acknowledge the support and input by members of the ASU Biodiversity Knowledge Center and ASU School of Life Sciences.

References

Allen C, Mehler DMA. 2019. Open science challenges, benefits and tips in early career and beyond. PLoS Biology 17(5): e3000246. https://doi.org/10.1371/journal.pbio.3000587

Ayres E. 2019. Quantitative guidelines for establishing and operating soil archives. Soil Science Society of America Journal 83: 973–981. https://doi.org/10.2136/sssaj2019.02.0050

Baker M. 2012. Building better biobanks. Nature 486: 141–146. https://doi.org/10.1038/486141a

Bálint M, Domisch S, Engelhardt CHM, Haase P, Lehrian S, Sauer J, Theissinger K, Pauls SU, Nowak S. 2011. Cryptic biodiversity loss linked to global climate change. Nature Climate Change 1: 313–318. https://doi.org/10.1038/nclimate1191

Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R. 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. Proceedings of the National Academy of Sciences 110: 4656–4660. https://doi.org/10.1073/pnas.1218503110

Buchholz S, Kreuels M, Kronshage A, Terlutter H, Finch O-D. 2011. Bycatches of ecological field studies: Bothersome or valuable? Methods in Ecology and Evolution 2: 99–102. https://doi.org/10.1111/j.2041-210X.2010.00051.x

Caballero A, García-Dorado A. 2013. Allelic diversity and its implications for the rate of adaptation. Genetics 195: 1373–1384. https://doi.org/10.1534/genetics.113.158410

Cash DW, Borck JC, Patt AG. 2006. Countering the loading-dock approach to linking scienceand decision making: Comparative analysis of El Niño/Southern Oscillation (ENSO) forecasting systems. Science, Technology, & Human Values 31: 465–494. https://doi.org/10.1177/0162243906287547

Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017. Going deeper in the automated identification of herbarium specimens. BMC Evolutionary Biology 17: 181. https://doi.org/10.1186/s12862-017-1014-z

Cook JA, Galbreath KE, Bell KC, Campbell ML, Carrière S, Colella JP, Dawson NG, Dunnum JL, Eckerlin RP, Greiman SE, et al. 2017. The Beringian coevolution project: Holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. Arctic Science 3: 585–617. https://doi.org/10.1139/as-2016-0042

Curty RG, Crowston K, Specht A, Grant BW, Dalton ED. 2017. Attitudes and norms affecting scientists' data reuse. PLoS ONE 12(12): e0189288. https://doi.org/10.1371/journal.pone.0189288

Dalton R. 1999. Monitoring system planned for US biodiversity drive. Nature 398: 738.

Damalas D, Kalyvioti G, Sabatella EC, Stergiou KI. 2018. Open data in the life sciences: The 'Selfish Scientist Paradox'. Ethics in Science and Environmental Politics 18: 27–36. DOI: 10.3354/esep00182

Daru BH, Park DSS, Primack RB, Willis CG, Barrington DS, Whitfield TJS, Seidler TG, Sweeney PW, Foster DR, Ellison AM, Davis CC. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytologist 217: 939–955. https://doi.org/10.1111/nph.14855

de Paula GA, Melo GA. 2015. Inferring sex and caste seasonality patterns in three species of bumblebees from Southern Brazil using biological collections. Neotropical Entomology 44: 10–20. DOI: 10.1007/s13744-014-0251-8

Dietze MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, Keitt TH, Kenney MA, Laney CM, Larsen LG, et al. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences of the United States of America 115: 1424–1432. https://doi.org/10.1073/pnas.1710231115

Edwards JL. 2004. Research and societal benefits of the Global Biodiversity Information Facility. BioScience 54: 485–486. https://doi.org/10.1641/0006 3568(2004)054[0486:RASBOT]2.0.CO;2

Escalas A, Hale L, Voordeckers JW, Yang Y, Firestone MK, Alvarez-Cohen L, Zhou J. 2019. Microbial functional diversity: From concepts to applications. Ecology and Evolution 9: 12000–12016. https://doi.org/10.1002/ece3.5670

Fernández-Triana J, Smith AM, Boudreault C, Goulet H, Hebert P, Smith AC, Roughley R. 2011. A poorly known high-latitude parasitoid wasp community: Unexpected diversity and dramatic changes through time. PLoS ONE 6(8): e23719.

Fontaine B, Perrard A, Bouchet P. 2012. 21 years of shelf life between discovery and description of new species. Current Biology 22: R943–R944.

Fontaine C, Dajoz I, Meriguet J, Loreau M. 2006. Functional diversity of plant-pollinator interaction webs enhances the persistence of plant communities. PLoS Biology 4(1): e1. https://doi.org/10.1016/j.cub.2012.10.029

Franz NM, Sterner BW. 2018. To increase trust, change the social design behind aggregated biodiversity data. Database, Volume 2018, 01 January 2018: bax100. https://doi.org/10.1093/database/bax100

Frey JK. 2009. Distinguishing range expansions from previously undocumented populations using background data from museum records. Diversity and Distributions 15: 183–187. https://doi.org/10.1111/j.1472-4642.2008.00552.x

Fullerton SM, Anderson NR, Guzauskas G, Freeman D, Fryer-Edwards K. 2010. Meeting the governance challenges of next-generation biorepository research. Science Translational Medicine 2(15): 15cm3. DOI: 10.1126/scitranslmed.3000361 DOI: 10.1126/scitranslmed.3000361

Garcillan PP, Ezcurra E. 2011. Sampling procedures and species estimation: Testing the effectiveness of herbarium data against vegetation sampling in an oceanic island. Journal of Vegetation Science 22: 273–280. https://doi.org/10.1111/j.1654-1103.2010.01247.x

Gibson, Cara M., Rebecca H. Kao, Kali K. Blevins, Patrick D. Travers. 2012. Integrative Taxonomy for Continental-Scale Terrestrial Insect Observations. PLOS ONE 7 (5): e37528. https://doi.org/10.1371/journal.pone.0037528.

Gomes VH, IJff SD, Raes N, Amaral IL, Salomão RP, de Souza Coelho L, de Almeida Matos FD, Castilho CV, de Andrade Lima Filho D, Cárdenas López D, et al. 2018. Species distribution modelling: Contrasting presence-only models with plot abundance data. Scientific Reports 8: 1003. https://doi.org/10.1038/s41598-017-18927-1

Gries C, Gilbert EE, Franz NM. 2014. Symbiota – a virtual platform for creating voucher-based biodiversity informatics communities. Biodiversity Data Journal 2: e1114.

Guralnick RP, Cellinese N, Deck J, Pyle RL, Kunze J, Penev L, Walls R, Hagedorn G, Agosti D, Wieczorek J, Catapano R, Page RDM. 2015. Community next steps for making Globally Unique Identifiers work for biocollections data. ZooKeys 494: 133–154. doi: 10.3897/zookeys.494.9352

Heberling JM, Isaac BL. 2017. Herbarium specimens as exaptations: New uses for old collections. American Journal of Botany 104: 963–965. DOI: 10.3732/ajb.1700125

Heberling JM, Prather LA, Tonsor SJ. 2019. The changing uses of herbarium data in an era of global change: An overview using automated content analysis. BioScience 69: 812–822. https://doi.org/10.1093/biosci/biz094

Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC, Edwards SV, Davis CC. 2020. Digitization and the future of natural history collections. BioScience 70: 243–251. https://doi.org/10.1093/biosci/biz163

Hobbie JE, Carpenter SR, Grimm NB, et al. 2003. The US Long Term Ecological Research Program. BioScience 53: 21–32. https://doi.org/10.1641/0006-3568(2003)053[0021:TULTER]2.0.CO;2

Hobern D, Baptiste B, Copas K, Guralnick R, Hahn A, van Huis E, Kim E-S, McGeoch M, Naicker I, Navarro L, et al. 2019. Connecting data and expertise: A new alliance for biodiversity knowledge. Biodiversity Data Journal 7: e33679.

Hoekman D, LeVan KE, Gibson C, Ball GE, Browne RA, Davidson RL, Erwin TL, Knisley CB, LaBonte JR, Lundgren J, et al. 2017. Design for ground beetle abundance and diversity sampling within the National Ecological Observatory Network. Ecosphere 8: e01744. https://doi.org/10.3897/BDJ.7.e33679

Holmes MW, Hammond TT, Wogan, GOU, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM, Nachman MW. 2016. Natural history collections as windows on evolutionary processes. Molecular Ecology 25: 864–881. DOI: 10.1111/mec.13529

Jeppsson T, Lindhe A, Gärdenfors U, Forslund P. 2010. The use of historical collections to estimate population trends: A case study using Swedish longhorn beetles (Coleoptera: Cerambycidae). Biological Conservation 143: 1940–1950. https://doi.org/10.1016/j.biocon.2010.04.015

Jetz W, McGeoch MA, Guralnick, Ferrier S, Beck J, Costello MJ, Fernandez M, Geller GN, Keil P, Merox C., et al. 2019. Essential biodiversity variables for mapping and monitoring species populations. Nature Ecology & Evolution 3: 539–551. https://doi.org/10.1038/s41559-019-0826-1

Johnson KG, Brooks SJ, Fenberg PB, Glover AG, James KE, Lister AM, Michel E, Spencer M, Todd JA, Valsami-Jones E, Young JR, Stewart JR. 2011. Climate change and biosphere response: Unlocking the collections vault. BioScience 61: 147–153. https://doi.org/10.1525/bio.2011.61.2.10

Johnston MA, Aalbu RL, Franz NM. 2018. An updated checklist of the Tenebrionidae sec. Bousquet et al. 2018 of the Algodones Dunes of California, with comments on checklist data practices. Biodiversity Data Journal 6: e24927.

Jones CA, Daehler CC. 2018. Herbarium specimens can reveal impacts of climate change on plant phenology; a review of methods and applications. PeerJ 6: e4576.

Kampe TU, McCorkel J, Hamlin L, Green RO, Krause KS, Johnson BR. 2011. Progress in the development of airborne remote sensing instrumentation for the National Ecological Observatory Network. Proceedings of the SPIE Optical Engineering + Applications, 81560A, Remote Sensing and Modeling of Ecosystems for Sustainability VIII (15 September 2011); https://doi.org/10.1117/12.892572

Kao RH, Gibson CM, Gallery RE, Meier CL, Barnett DT, Docherty KM, Blevins KK, Travers PD, Azuaje E, Springer YP, Thibault KM, McKenzie VJ, Keller M, Alves LF, Hinckley E-LS, Parnell J, Schimel D. 2012. NEON terrestrial field observations: Designing continental-scale, standardized sampling. Ecosphere 3(12): 115. https://doi.org/10.1890/ES12-00196.

Karan M, Liddell M, Prober SM, Arndt S, Beringer J, Boer M, Cleverly J, Eamus D, Grace P, Van Gorsel E, Hero J-M, Hutley L, Macfarlane C, Metcalfe D, Meyer W, Pendall E, Sebastian A, Wardlaw T. 2016. The Australian SuperSite Network: A continental, long-term terrestrial ecosystem observatory. Science of the Total Environment 568: 1263–1274. https://doi.org/10.1016/j.scitotenv.2016.05.170

Keller M, Schimel DS, Hargrove WW, Hoffman FM. 2008. A continental strategy for the National Ecological Observatory Network. Frontiers in Ecology and the Environment 6: 282–284. https://doi.org/10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2

Kharouba HM, Lewthwaite JMM, Guralnick R, Kerr JT, Vellend M. 2018. Using insect natural history collections to study global change impacts: Challenges and opportunities. Philosophical Transactions of the Royal Society B, Biological Science, 374: 20170405. https://doi.org/10.1098/rstb.2017.0405

Kissling WD, Walls R, Bowser A, Jones MO, Kattge J, Agosti D, Amengual J, Basset A, van Bodegom PM, Cornelissen JHC, et al. 2018. Towards global data products of Essential Biodiversity Variables on species traits. Nature Ecology and Evolution 2: 1531–1540. https://doi.org/10.1038/s41559-018-0667-3

Larson ER, Graham BM, Achury R, Coon JJ, Daniels MK, Gambrell DK, Jonasen KL, King GD, LaRacuente N, Perrin-Stowe TIN, et al. 2020. From eDNA to citizen science: Emerging tools for the early detection of invasive species. Frontiers in Ecology and the Environment. https://doi.org/10.1002/fee.2162

Laurance WF, Nascimento HE, Laurance SG, Andrade A, Ribeiro JE, Giraldo JP, Lovejoy TE, Condit R, Chave J, Harms KE, D'Angelo S. 2006. Rapid decay of tree-community composition in Amazonian forest fragments. Proceedings of the National Academy of Sciences of the United States of America 103: 19010–19014. doi: 10.1073/pnas.0609048103.

Lemos MC, Arnott JC, Ardoin NM, Baja K, Bednarek AT, Dewulf A, Fieseler C, Goodrich KA, Jagannathan K, Klenk N. 2018. To co-produce or not to co-produce. Nature Sustainability 1: 722–724. https://doi.org/10.1038/s41893-018-0191-0

Lendemeer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC. 2020. The Extended Specimen Network: A strategy to enhance US biodiversity collections, promote research and education. BioScience 70: 23–30. https://doi.org/10.1093/biosci/biz140

Leonelli S. 2019. What distinguishes data from models? European Journal for Philosophy of Science 9: 22. https://doi.org/10.1007/s13194-018-0246-0

Lewis JS, Farnsworth ML, Burdett CL, Theobald DM, Gray M, Miller RS. 2017. Biotic and abiotic factors predicting the global distribution and population density of an invasive large mammal. Scientific Reports 7: 44152. https://doi.org/10.1038/srep44152

Linard B, Arribas P, Andújar C, Crampton-Platt A, Vogler AP. 2016. Lessons from genome skimming of arthropod-preserving ethanol. Molecular Ecology Resources 16: 1365 -1377. https://doi.org/10.1111/1755-0998.12539

Loranger J, Shipley B. 2010. Interspecific covariation between stomatal density and other functional leaf traits in a local flora. 2010. Botany. 88: 30–38. https://doi.org/10.1139/B09-103

Lütolf M, Kienast F, Guisan A. 2006. The ghost of past species occurrence: Improving species distribution models for presence-only data. Journal of Applied Ecology 43: 802–815. DOI: 10.1111/j.1365-2664.2006.01191.x

Lynggaard C, Nielsen M, Santos-Bay L, Gastauer M, Oliveira G, Bohmann K. 2019. Vertebrate Diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. Environmental DNA 1: 329–341.m https://doi.org/10.1002/edn3.34

McLauchlan KK, Ferguson CJ, Wilson IE, Ocheltree TW, Craine JM. 2010. Thirteen decades of foliar isotopes indicate declining nitrogen availability in central North American grasslands. New Phytologist 187: 1135–1145. https://doi.org/10.1111/j.1469-8137.2010.03322.x

McLean BS, Bell KC, Dunnum JL, Abrahamson B, Colella JP, Deardorff ER, Weber JA, Jones AK, Salazar-Miralles F, Cook JA. 2016. Natural history collections-based research: Progress, promise, and best practices. Journal of Mammalogy 97: 287–297. https://doi.org/10.1093/jmammal/gyv178

Meineke EK, Classen AT, Sanders NJ, Jonathan Davies T. 2019. Herbarium specimens reveal increasing herbivory over the past century. Journal of Ecology 107: 105–117. https://doi.org/10.1111/1365-2745.13057

Meineke EK, Davies TJ, Daru BH, Davis CC. 2018. Biological collections for understanding biodiversity in the Anthropocene. Philosophical Transactions of the Royal Society B, Biological Sciences, 374: 20170386. https://doi.org/10.1098/rstb.2017.0386

Meyer C, Weigelt P, Kreft H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters 19: 992–1006. NEON – National Ecological Observatory Network. 2019. doi: 10.1111/ele.12624.

Penuelas J, Jannssens IA, Ciais P, Obersteiner M, Sardans J. 2020. Anthropogenic global shifts in biospheric N and P concentrations and ratios and their impacts on biodiversity, ecosystem productivity, food security, and human health. Global Change Biology 26: 1962–1985. https://doi.org/10.1111/gcb.14981

Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG, Scholes RJ, Bruford MW, Brummitt N, Butchart SHM, Cardoso AC, Coops NC, Dulloo E, Faith DP, Freyhof J, Gregory RD, Heip C, Höft R, Hurtt G, Jetz W, Karp DS, McGeoch MA, Obura D, Onoda Y, Pettorelli N, Reyers B, Sayre R, Scharlemann JPW, Stuart SN, Turak E, Walpole M, Wegmann M. 2013. Essential biodiversity variables. Science 339: 277–278. DOI: 10.1126/science.1229931

Pérez-Ramos IM, Díaz-Delgado R, de la Riva EG, Villar R, Lloret F, Marañón T. 2017. Climate variability and community stability in Mediterranean shrublands: The role of functional diversity and soil environment. Journal of Ecology 105: 1335–1346. https://doi.org/10.1111/1365-2745.12747

Ponder WF, Carter GA, Flemons P, Chapman RR. 2001. Evaluation of museum collection data for use in biodiversity assessment. Conservation Biology 15: 648–657. https://doi.org/10.1046/j.1523-1739.2001.015003648.x

Pyke GH, Ehrlich PR. 2010. Biological collections and ecological/environmental research: A review, some observations and a look to the future. Biological Reviews 85: 247–266. https://doi.org/10.1111/j.1469-185X.2009.00098.x

Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi K, Moritz CC. 2011. Museum genomics: Low-cost and high-accuracy genetic data from historical specimens. Molecular Ecology Resources 11: 1082–1092. https://doi.org/10.1111/j.1755-0998.2011.03052.x

Rudin SM, Murray DW, Whitfeld TJS. 2017. Retrospective analysis of heavy metal contamination in Rhode Island based on old and new herbarium specimens. Applications in Plant Sciences 5(1): 1600108. doi: 10.3732/apps.1600108

SanClements M, Lee RH, Ayres ED, Goodman K, Jones M, Durden D, Thibault K, Zulueta R, Roberti J, Lunch C, Gallo A. 2020. Collaborating with NEON. BioScience 70: 107. https://doi.org/10.1093/biosci/biaa005

Schilthuizen M, Vairappan CS, Slade EM, Mann DJ, Miller JA. 2015. Specimens as primary data: Museums and 'open science'. Trends in Ecology and Evolution 30: 237–238. https://doi.org/10.1016/j.tree.2015.03.002

Schindel DE, Cook JA. 2018. The next generation of natural history collections. PLoS Biology 16: e2006125.

Scholes RJ, Walters M, Turak E, Saarenmaa H, Heip CHR, Tuama ÉÓ, Faith DP, Mooney HA, Ferrier S, Jongman RHG. 2012. Building a global observing system for biodiversity.

Current Opinion in Environmental Sustainability 4: 139–146. https://doi.org/10.1016/j.cosust.2011.12.005

Senkowsky S. 2003. NEON: Planning for a new frontier in biology. BioScience 53: 456–461. https://doi.org/10.1641/0006-3568(2003)053[0456:NPFANF]2.0.CO;2

Sharma S, Ciufo S, Starchenko E, Darji D, Chlumsky L, Karsch-Mizrachi I, Schoch CL. 2018. The NCBI BioCollections Database. Database 2018: bay006. doi: 10.1093/database/bay006

Spears LR, Ramirez R. 2015. Learning to love leftovers. American Entomologist 61: 168–173. https://doi.org/10.1093/ae/tmv046

Strasser BJ. 2019. Collecting Experiments: Making Big Data Biology. Oxford University Press. DOI: 10.7208/chicago/9780226635187.001.0001

Suarez AV, Tsutsui ND. 2004. The value of museum collections for research and society. BioScience 54: 66–74. https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2

Taylor BW, Flecker AS, Hall Jr RO. 2006. Loss of a harvested fish species disrupts carbon flow in a diverse tropical river. Science 313(5788): 833–836. DOI: 10.1126/science.1128223

Thackeray SJ, Henrys PA, Hemming D, Bell JR, Botham MS, Burthe S, Helaouet P, Johns DG, Jones ID, Leech DI, et al. 2016. Phenological sensitivity to climate across taxa and trophic levels. Nature 535: 241–245. https://doi.org/10.1038/nature18608

Thorpe AS, Barnett DT, Elmendorf SC, Hinckley ELS, Hoekman D, Jones KD, LeVan KE, Meier CL, Stanish LF, Thibault KM. 2016. Introduction to the sampling designs of the National Ecological Observatory Network Terrestrial Observation System. Ecosphere 7(12): e01627. https://doi.org/10.1002/ecs2.1627

Tylianakis JM, Laliberté E, Nielsen A, Bascompte J. 2010. Conservation of species interaction networks. Biological Conservation 143: 2270–2279. https://doi.org/10.1016/j.biocon.2009.12.004

Van Allen A. 2018. Pinning beetles, biobanking futures: Practices of archiving life in a time of extinction. New Genetics and Society 37: 387–410. DOI: 10.1080/14636778.2018.1546573

W3C, World Wide Web Consortium. 2017. Data on the Web Best Practices. https://www.w3.org/TR/2017/REC-dwbp-20170131/

Waide RB, Brunt JW, Servilla MS. 2017. Demystifying the landscape of ecological data repositories in the United States. BioScience 67: 1044–1051. https://doi.org/10.1093/biosci/bix117

Ward DF. 2012. More than just records: Analysing natural history collections for biodiversity planning. PLoS ONE 7(11): e50346.

Ward DF, Leschen RAB, Buckley TR. 2015. More from ecologists to support natural history museums. Trends in Ecology and Evolution 30: 373–374. DOI: 10.1016/j.tree.2015.04.015

Webster MS (editor). 2017. The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research. Studies in Avian Biology Series, No. 50. CRC press.

Weeks BC, Willard DE, Zimova M, Ellis AA, Witynski ML., Hennen M, Winger BM. 2019. Shared morphological consequences of global warming in North American migratory birds. Ecology Letters 23: 316–325. https://doi.org/10.1111/ele.13434

Wenger E. 2000. Communities of practice and social learning systems. Organization 7: 225–246. https://doi.org/10.1177/135050840072002

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7(1): e29715.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, Bonino de Silva Santos L, Bourne PE. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18

Willing E-M, Dreyer C, van Oosterhout C. 2012. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. PLoS ONE. 7(8): e42649. doi: 10.1371/journal.pone.0042649

Wilson, N. 2020. Medical Entomologists: A Vanishing Species: Experts needed to combat vector-borne diseases in North America. BioScience. 70(4): 281–288. https://doi.org/10.1093/biosci/biaa008

Yoshida K, Burbano HA, Krause J, Thines M, Weigel D, Kamoun S. 2014. Mining herbaria for plant pathogen genomes: Back to the future. PLoS Pathogens 10(4): e1004028.

Zhang Q, Buyantuev A, Li FY, Jiang L, Niu J, Ding Y, Kang S, Ma W. 2017. Functional dominance rather than taxonomic diversity and functional diversity mainly affects community aboveground biomass in the Inner Mongolia grassland. Ecology and Evolution 7: 1605–1615. https://doi.org/10.1002/ece3.2778

Ziska LH, Pettis JS, Edwards J, Hancock JE, Tomecek MB, Clark A, Dukes JS, Loladze I, Polley HW. 2016. Rising atmospheric CO2 is reducing the protein concentration of a floral pollen source essential for North American bees. Proceedings of the Royal Society B, Biological Sciences, 283: 20160414. DOI: 10.1098/rspb.2016.0414

Zulkafli Z, Perez K, Vitolo C, Buytaert W, Karpouzoglou T, Dewulf A, De Bièvre B, Clark J, Hannah DM, Shaheed S. 2017. User-driven design of decision support systems for polycentric environmental resources management. Environmental Modelling & Software 88: 58–73. DOI: 10.1016/j.envsoft.2016.10.012

Box 1. Glossary.

Biobank. Biocollection that focuses on preserving samples – including living cells, tissues, dried and frozen samples, etc. – primarily for the biomedical (human health) research domain (Baker 2012).

Biocollection. Any collection preserving material organismal and environmental samples for research purposes.

Biorepository. In the biomedical field, this term is sometimes used interchangeably with biobank (e.g. Brigham and Women's Hospital, Johns Hopkins Pediatrics and the University of California San Francisco biorepositories). Here, we use it to describe biocollections with a focus on biodiverse organismal and environmental samples, with an explicit, constrained purpose to optimize sample composition and processing to monitor varied responses to ecological and evolutionary change.

Darwin Core standard. Data standard for publishing and integrating biodiversity information (Wieczorek et al. 2012). Darwin Core is widely used in natural history collections and observational communities and projects that focus on documenting and monitoring biodiversity; including the Global Biodiversity Information Facility.

Material sample. Physical sample – ranging from bulk (multi-specimen) samples, individual specimens, tissues, DNA, to environmental samples – preserved in biocollections.

Natural history collection. Biocollection developed primarily in the tradition of discovering and documenting (macroorganismal) biodiversity regionally and/or globally.

Observation. Recording of an occurrence (in the sense of Darwin Core) without material sample preservation; e.g. promoted by projects such as eBird, iNaturalist, among many others.

Specimen. Material sample that corresponds to an individual (macro)organism, often prepared separately from other such specimens preserved in biocollections.

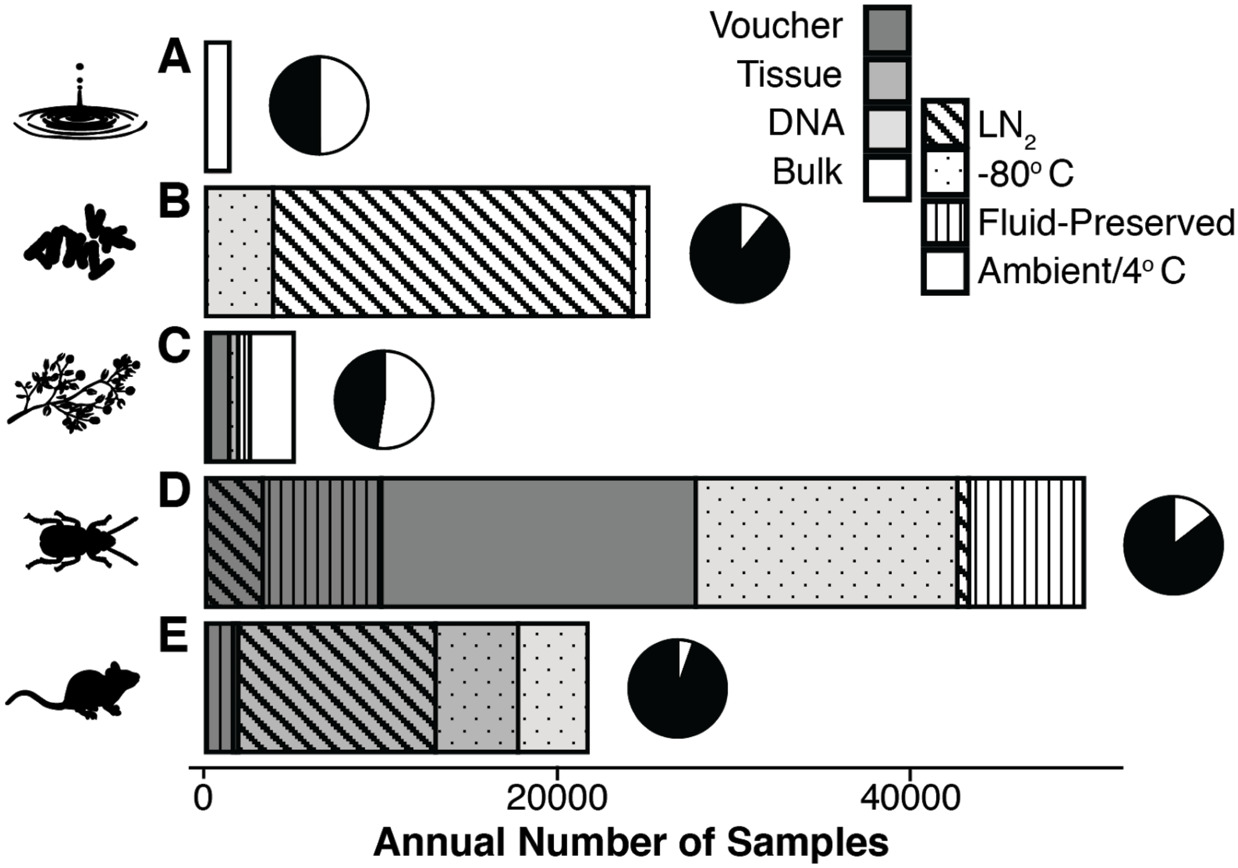

Figure 1. Estimated annual NEON Biorepository sample intake. Samples of the following taxonomic categories are shown: (A) environmental; (B) aquatic and soil microbe; (C) algal, lichen, bryophyte, and terrestrial plant; (D) invertebrate; and (D) vertebrate. Shading of the bars indicates sample type: whole-organism vouchers, including single species lots; tissue (e.g. blood, hair, fecal, or leaf) sample; DNA sample; or bulk sample containing many organisms of multiple species. Bar patterning indicates sample storage conditions: “LN_2” is stored in cryo-safe containers in liquid nitrogen; “-80℃” is stored in ultralow mechanical freezer; “Fluid-Preserved” is stored in either ethanol, glutaraldehyde, or lugols; and “Ambient/4℃” is stored dry, slide-mounted, or mechanical refrigerator. Pie charts associated with each taxon represent the proportions of samples originating from the NEON aquatic sites (white) and terrestrial sites (black).

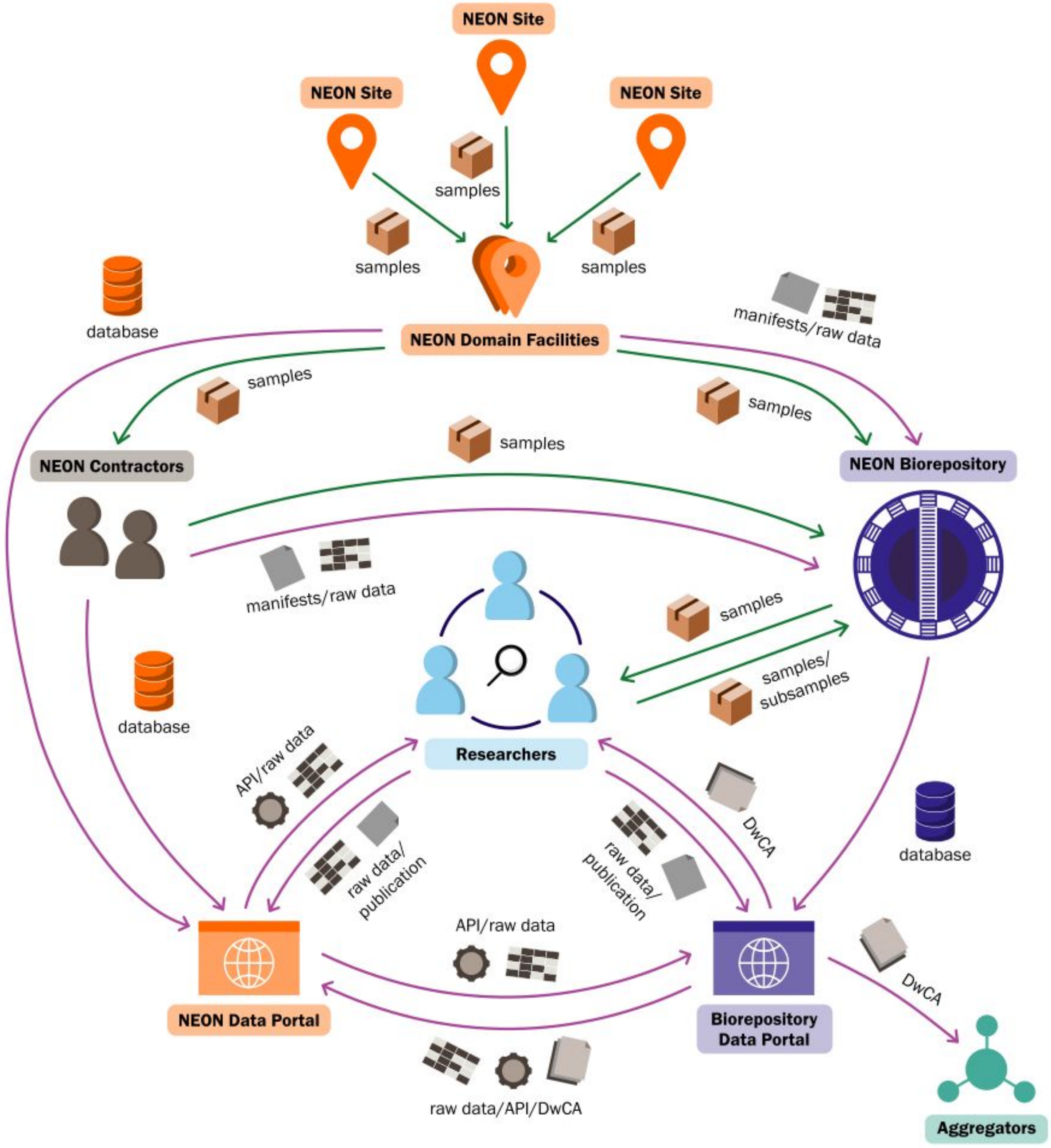

Figure 2. Sample and data flow diagram showing interactions between different actors and repositories to be fully integrated with the NEON Biorepository. The NEON Biorepository should function like a turn table that dynamically assimilates, redistributes, and reintegrates samples and (value-added) data through innovative biodata science practices. Green arrows represent sample transfer processes, whereas purple arrows represent data transfer or transformation processes. Lowering thresholds for engaging research teams (light blue) is the overarching

strategic goal. To achieve this, improving the consistency and efficiency in creating and propagating new data records and subsequent annotations throughout the entire internal NEON network, and through involvement of researchers, is critical.