Integrated use of enhanced natural history collections is key to solve the biodiversity crisis

Julia D. Sigwart*[1], Matthias Schleuning*[2], Angelika Brandt[1,3], Markus Pfenninger[2,4], Senckenberg Collectomics Consortium

these authors contributed equally

1. Senckenberg Research Institute and Natural History Museum, Department of Marine Zoology, Frankfurt am Main, Germany
2. Senckenberg Biodiversity Climate Research Center (SBiK-F), Frankfurt am Main, Germany
3. Institute for Ecology, Evolution and Diversity, Goethe-University of Frankfurt, Frankfurt am Main, Germany
4. LOEWE Centre for Translational Biodiversity Genomics, Frankfurt am Main, Germany

Address for correspondence: julia.sigwart@senckenberg.de

Introduction

Global biodiversity loss is arguably the biggest problem facing humanity. Climate change, changes in land and sea use and other factors are synergistically eroding biodiversity to an unprecedented speed and extent, with cascading impacts on humanity and our livelihoods (1). Scientific advice on safeguarding biodiversity depends on all available information to understand past and current developments, and predict future responses of Earth's ecosystems. This challenge requires integrative research across space, time, methods, and taxa, and integration of these data into a new generation of biodiversity models (2, 3). Such research is currently thwarted because biodiversity data are stored in different formats and databases (4, 5), and the largest sources of biodiversity data are still contained in physical repositories that are not fully accessible: collections of geological and biological specimens (Figure 1). To overcome this shortfall, natural history collections must be developed into specimen-based, integrated, and digitally accessible research platforms. We propose that a new conceptual framework, *Collectomics,* is required to underpin this vision; this aegis embraces the entirety of collection-based research, although we focus here on how it enables and fuels the research necessary to effectively confront the Anthropocene biodiversity crisis (Figure 2). Current technological developments provide an unprecedented opportunity to unleash the full potential of collections by fully integrating the myriad data dimensions from collection objects.

The extended natural history collection

Natural history collections comprise a vast, de-centralised, continuously growing global infrastructure of biodiversity data, central to taxonomy, phylogenetics, biogeography and Earth system research (6); scientific specimens with their associated data and archives contribute to effectively all areas of geo- and biosciences. Global collections are an irreplaceable archive preserving data continuously from the beginning of the Anthropocene and long before. This encompasses a much longer time scale than the decade-scale series usually captured in biodiversity monitoring programs (7). As such, collections are increasingly recognised as the most important data source to provide comparative baselines on historical time scales and in deep time (8,9). Recent calls to increase attention to the value of collections have focussed on their potential for critical areas such as drug discovery or emergent diseases (10); yet the potential value of collections remains largely untapped even in modern biodiversity research, because the vast scope of collections dwarfs current efforts at data mobilisation.

Mobilising global metadata is often the focus of collections' contributions to biodiversity (6). In parallel, developments in biodiversity and ecosystem research revolve around the ambition to bridge multiple scales from species-level to ecosystem-level perspectives, by using trait-based approaches and models (5, 11). Specimens in collections preserve trait data across different realms, taxonomic groups and temporal and spatial scales, yet recent developments in ecosystem research have been largely disconnected from collection-based research.

The real added value of collection-based data is therefore not only large-scale occurrence data over time and space, but that the information is tied to physical specimens (12, 13). Type specimens are the definitive authority for taxonomic species descriptions, increasingly in demand for the global quest to describe extant and extinct life on earth and conserve living species diversity (9, 14, 15). Biodiversity studies at small and large scales are increasingly hampered by cascading errors of species identification, especially as cryptic species are revealed by DNA data (16). The preservation of physical voucher specimens underpins replicability of all collections-based research, including species discovery, and ensures the interoperability of data across different data dimensions (6, 17). Painstakingly assembled datasets from collections have revealed fundamental mechanisms of Earth-system dynamics, such as mass extinctions and tipping points, latitudinal diversity gradients, and evolution through escalation (18, 19). The next generation of collection-based approaches share the focus on specimens as a physical database, but with the promise of accelerating and expanding access (15).

The conceptual extension from a physical specimen to its structural and functional properties mirrors the development in ecosystem research, moving away from counting species to measuring how species interact with their environment and contribute to ecosystem functioning (11). Museum specimens physically preserve a wealth of data, including genomes, morphological traits, environmental signatures, and many more dimensions that inform the understanding of ecological and evolutionary processes and ecosystem functions. The potential of the "extended specimen" (13) is increasingly unlocked with rapidly advancing technologies, such as biogeochemical methods, genome sequencing, hyperspectral imaging, or X-ray tomography, as well as future advances that cannot be imagined now (6,12, 20). Integrating multiple data dimensions of the extended specimen will enable research centred on different levels of organization (e.g., genes, traits, functions, ecosystems, Earth system) and might bridge geo- and biosciences, as well as genomics, ecological and evolutionary research (21). Ultimately, this might enable multi-scale approaches to understand biodiversity dynamics and the consequences of biodiversity loss, from genetic diversity over trait diversity to changes in ecosystems and the Earth system (22). This new generation of trait-based research remains indivisibly tied to taxonomic research, linking trait datasets with species' geographical distribution and phylogenetic data.

Future perspectives and challenges

Unleashing the potential of collections to support biodiversity in the Anthropocene depends on increasing accessibility of collections data, their integration across disciplines, and ensuring the Findability, Accessibility, Interoperability, and Reusability (FAIR) for past, present and future collection objects and data (6). This is possible now, as tremendous developments in digitisation and the build-up of global research networks offer unprecedented opportunities to develop collections into integrated research platforms (23). Rapidly developing applications from artificial intelligence and machine learning finally promise realistic major advances in mobilising specimen data (20, 24), as well as the long-anticipated applications for computer-aided species identification, and trait recognition (20, 25). Ongoing mass digitization of the temporal and spatial data attached to physical specimens will provide the main reference for large-scale and long-term studies of biodiversity. What is still needed are fully interoperable repositories of genomes, traits and other data that can bridge the two ends of the spectrum: from detailed insights from single specimens to large-scale analyses of ecosystems (4, 5).

The aim of biodiversity research reaches beyond describing patterns and processes under current conditions, to understanding ecosystem changes in the context of past, ongoing and future global change. Species-specific approaches fall short in describing the diversity of global ecosystems, but can be generalized with trait-based approaches that describe the functional diversity of life on Earth (26). For example, trait-based approaches have enabled important insights into ecosystem resilience to climate change based on dynamic vegetation models (22). Such models started with a relatively coarse classification of plant diversity into a few functional types, but recent developments can actually incorporate the full variability in plant form and function into mechanistic and predictive modelling frameworks (27). Further advances in biodiversity models will require cross-taxon approaches that go beyond a plant-centred understanding of ecosystems and the integration of trait data from across taxa and trophic levels (3). While animal trait data are notoriously difficult to measure in the wild, collections provide a huge and largely untapped resource for compiling such datasets and using them in biodiversity models.

One challenge to our vision of integrating specimen-based and ecosystem-based research is the integration of specimen-based data across collections, taxa and research disciplines (5). Established repositories for each of these particular kinds of data exist (e.g., occurrence records, genetic sequences, phylogenetic trees, morphological or physiological traits), but they are rarely interoperable. While there are established global data standards for collections, with unique specimen-based permanent identifiers (28), current practices do not require other data repositories to connect to these standard specimen identifiers. Our envisioned integrative research agenda requires global initiatives of data sharing and integration, aiming at decentralized, but highly connected and interoperable networks of biodiversity data. This conceptual framework provided by Collectomics is intended to support this integration and emphasize the centrality of specimen-based data.

Collectomics

Collectomics is not a new way to use collections, it is a succinct term to rally efforts around the many facets of science that could benefit from collections-based research. It aspires to increase the integration with other types of relevant data, such as those derived from remote-sensing or large citizen-science initiatives (29). Global data mobilization of the authoritative data source for biodiversity — natural history collections — is further important for equitable accessibility for science and society (14). In the context of Anthropocene biodiversity loss, resources and contextual data are most needed in tropical countries of the Global South (30); however, collections resources that underpin the understanding of global biodiversity, especially type material, are mainly in Europe and North America. Accelerating mass digitization has a vast potential to make collections data available where they are actually needed.

Overcoming the shortfalls that impede global solutions to the biodiversity crisis requires all available resources, and is in urgent need of a new approach integrating collection-based and ecosystem-based research. We have identified two main levels where this integration will be essential: (i) describing biodiversity trends in space and time, and (ii) gaining a multidimensional understanding from genes through traits to ecosystems and Earth system dynamics. To unleash the full potential of natural history collections, we call for efforts to increase accessibility of collections data by establishing globally standardised, interoperable, and extensible data models centred on the specimen identifier, to allow seamless integration with species-based repositories. This promotes approaches combining data from across different biological levels (e.g., genomes, individuals, populations) with other data dimensions from extant and extinct organisms, such as their biotic and abiotic environments, to build a new generation of biodiversity models.

The benefits of collections are manifestly not limited to biodiversity and Earth system research, given their importance to many realms of science as well as bio-inspired design, art, education and countless other channels. Collections represent key scientific infrastructure and some of the oldest continuously maintained and expanded scientific resources on Earth. Collectomics is a new framework that is fundamental to integrate the modern applications of collections, including their pivotal role in finding solutions to Anthropocene challenges.

References

1. S. Díaz, *et al.*, Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
2. A.-M. Waldvogel, *et al.*, Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters* **4**, 4–18 (2020).
3. M. Schleuning, *et al.*, Trait-Based Assessments of Climate-Change Impacts on Interacting Species. *Trends in Ecology & Evolution* **35**, 319–328 (2020).
4. W. D. Kissling, *et al.*, Towards global data products of Essential Biodiversity Variables on species traits. *Nat Ecol Evol* **2**, 1531–1540 (2018).
5. R. V. Gallagher, *et al.*, Open Science principles for accelerating trait-based science across the Tree of Life. *Nat Ecol Evol* **4**, 294–303 (2020).
6. S.E. Miller, L.N. Barrow, S.M. Ehlman, J.A. Goodheart, S.E. Greiman, H.L. Lutz, T.M. Misiewicz, S.M. Smith, M. Tan, C.J. Thawley, J.A. Cook, J.A. Building natural history collections for the twenty-first century and beyond. *BioScience*, **70**, 674–687 (2020).
7. M. Dornelas, *et al.*, BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecol Biogeogr* **27**, 760–786 (2018).
8. E. K. Meineke, T. J. Davies, B. H. Daru, C. C. Davis, Biological collections for understanding biodiversity in the Anthropocene. *Phil. Trans. R. Soc. B* **374**, 20170386 (2019).
9. L. Lopez, K. G. Turner, E. S. Bellis, J. R. Lasky, Genomics of natural history collections for understanding evolution in the wild. *Molecular Ecology Resources* **20**, 1153–1160 (2020).
10. D. DiEuliis, K. R. Johnson, S. S. Morse, D. E. Schindel, Specimen collections should have a much bigger role in infectious disease research and response. *PNAS* **113**, 4–7 (2016).
11. J. L. Funk, *et al.*, Revisiting the Holy Grail: using plant functional traits to understand ecological processes: Plant functional traits. *Biol Rev* **92**, 1156–1173 (2017).
12. L. Sumner-Rooney, J. D. Sigwart, Lazarus in the museum: resurrecting historic specimens through new technology. *Invertzool* **14**, 73–84 (2017).
13. M. S. Webster, *The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research* (CRC Press, 2017).
14. J. D. Sigwart, *What Species Mean: A User's Guide to the Units of Biodiversity* (CRC Press, 2018).
15. A. Miralles, *et al.*, Repositories for Taxonomic Data: Where We Are and What is Missing. *Systematic Biology* **69**, 1231–1253 (2020).
16. M. Pfenninger, K. Schwenk, Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* **7**, 121 (2007).
17. J. P. Colella, *et al.*, The Open-Specimen Movement. *BioScience*, biaa146 (2020).
18. G. J. Vermeij, *Evolution and Escalation: An Ecological History of Life* (Princeton University Press, 1993).
19. K. Roy, D. Jablonski, J. W. Valentine, G. Rosenberg, Marine latitudinal diversity gradients: Tests of causal hypotheses. *PNAS* **95**, 3699–3702 (1998).

20. T. T. Høye, *et al.*, Deep learning and computer vision will transform entomology. *Proc Natl Acad Sci USA* **118**, e2002545117 (2021).
21. S. F. Ryan, *et al.*, Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *PNAS* **115**, E2284–E2291 (2018).
22. B. Sakschewski, *et al.*, Resilience of Amazon forests emerges from plant trait diversity. *Nature Clim Change* **6**, 1032–1036 (2016).
23. B. P. Hedrick, *et al.*, Digitization and the Future of Natural History Collections. *BioScience* **70**, 243–251 (2020).
24. V. Blagoderov, I. J. Kitching, L. Livermore, T. J. Simonsen, V. S. Smith, No specimen left behind: industrial scale digitization of natural history collections. *Zookeys*, 133–146 (2012).
25. T. Cordier, *et al.*, Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources* **18**, 1381–1391 (2018).
26. D. Purves, J.P. Scharlemann, M. Harfoot, T. Newbold, D.P. Tittensor, J. Hutton, S. Emmott, Time to model all life on Earth. *Nature*, **493,** 295–297 (2013).
27. S. Scheiter, L. Langan, S. I. Higgins, Next-generation dynamic global vegetation models: learning from community ecology. *New Phytol* **198**, 957–969 (2013).
28. G. Nelson, P. Sweeney, E. Gilbert, Use of globally unique identifiers (GUIDs) to link herbarium specimen records to physical specimens. *Applications in Plant Sciences* **6**, e1027 (2018).
29. D. M. Spear, G. B. Pauly, K. Kaiser, Citizen Science as a Tool for Augmenting Museum Collection Data from Urban Areas. *Front. Ecol. Evol.* **5** (2017).
30. A. Lira-Noriega, J. Soberón, The relationship among biodiversity, governance, wealth, and scientific capacity at a country level: Disaggregation and prioritization. *Ambio* **44**, 391–400 (2015).

Figure 1. Natural history collections are iconic in biodiversity research and yet much of their potential impact remains untapped. Photograph by Sven Tränkner, Senckenberg Museum Frankfurt, Germany.

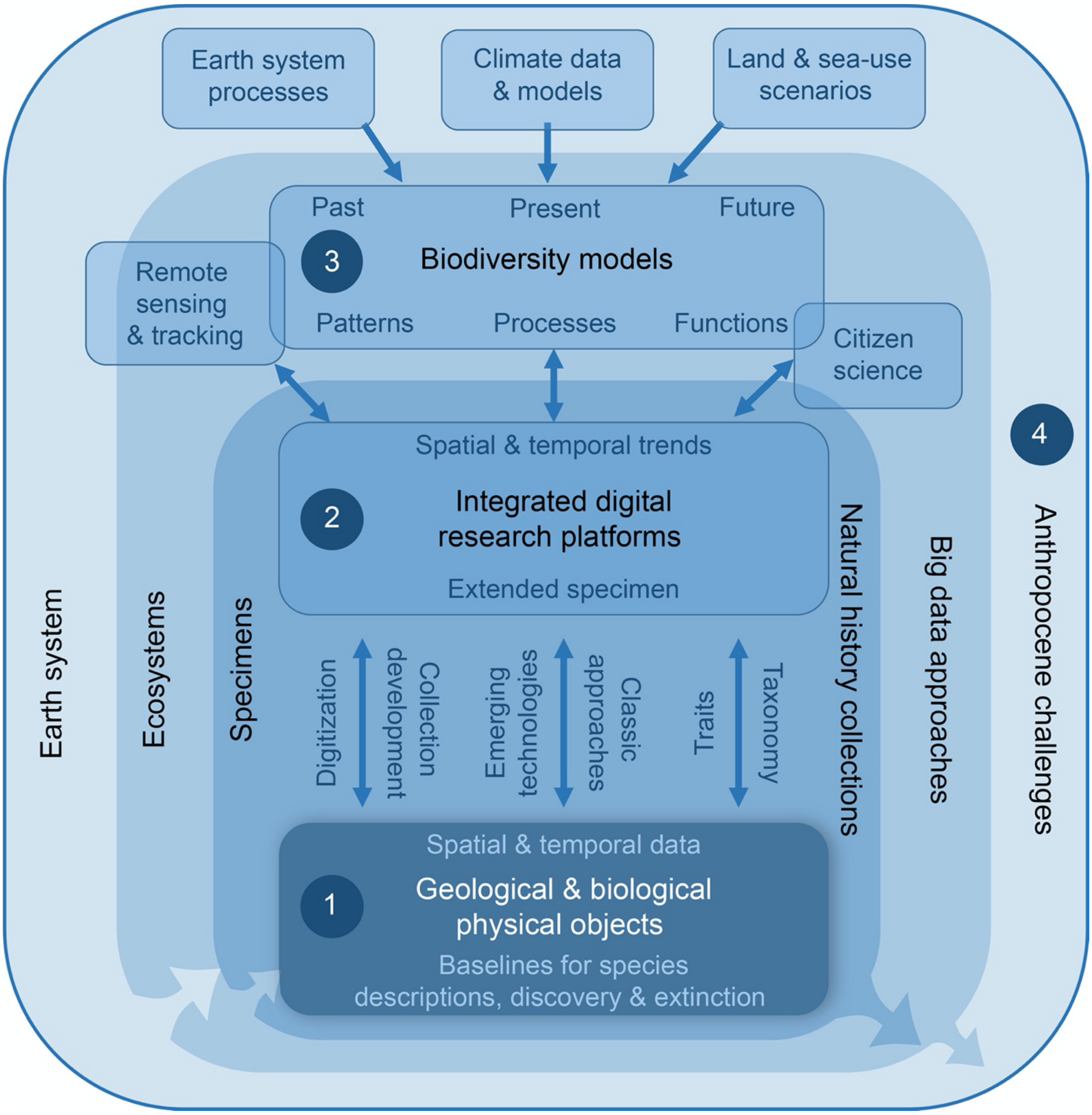

Figure 2. Collectomics describes the multiple ways collection-based research interacts synergistically with other research fields across scales (indicated by different shades of blue). Digital solutions can shift collections from decentralised physical repositories (**1**) towards fully integrated and interconnected research platforms (**2**), readily accessible to science and society across the globe. These platforms interact with other big-data approaches (e.g., remote sensing, citizen science) and are expected to underpin and inspire new developments in biodiversity research (**3**), e.g., in analyses of biodiversity trends across scales or process-based models of ecosystem functions. By linking this endeavour to data and approaches from Earth sciences, Collectomics can ultimately contribute to providing solutions to the Anthropocene challenges (**4**).