Biodiversity Science and the Twenty-First Century Workforce

ELIZABETH R. ELLWOOD, JOCELYN ANNE SESSA, JOEL K. ABRAHAM, AMBER E BUDDEN, NATALIE DOUGLAS, ROBERT GURALNICK, ERICA KRIMMEL, TOM LANGEN, DEBRA LINTON, MOLLY PHILLIPS, PAMELA S. SOLTIS, MARIE STUDER, LISA D. WHITE, JASON WILLIAMS, AND ANNA K. MONFILS

Amid growing global concern about biodiversity loss (IPBES 2019), scientists are conducting research to inventory life on Earth, develop new conservation and management strategies, and investigate emergent issues pertaining to invasive species spread, zoonotic disease transmission, and climate change impacts. As scientists seek to conserve biodiversity and understand the impacts of anthropogenic disturbance, natural history collections have proven to be a valuable resource. Natural history collections house specimens that document the taxonomy, evolution, cultural significance (as with zooarchaeological specimens), biogeography, and ecology of species (McLean et al. 2015). With over 2 billion physical specimens archived in natural history collections across the globe, scientists are able to access both the physical specimens and associated metadata (e.g., collection date, location, habitat, media, community assemblage; Marshall et al. 2018) that chronicle past and present biodiversity. National and international digitization efforts have begun to mobilize this wealth of collections data prompting the development of cyberinfrastructure capable of storing, accessing, and conducting novel research using these large data sets (such as those available via gbif.org, idigbio.org, neotomadb.org, paleobiodb.org). The physical specimen and its digital record provide a rich resource of genetic, phenotypic, behavioral, geographic, and environmental data ("the extended specimen"; Webster 2017, Biodiversity Collections Network 2019). When specimen-based data are linked to environmental data sets (e.g., WorldClim, PRISM), the research potential of the specimens is further enriched. As the data and resources are mobilized and integrated, scientists are conducting deeper and increasingly more complex analyses that model future impacts of climate change, research drivers of species loss, investigate phenological asynchrony, and more (Meineke et al. 2018 and the references therein).

The promise of integrative, data-rich, computational approaches for understanding and conserving biodiversity has never been greater or more attainable. However, many professionals and students alike lack the expertise needed to access and synthesize the gamut of resources needed for integrative analyses. Newly mobilized data have, in conjunction with the tools and skill sets necessary to access and analyze them, put a premium on the workforce skills and competencies needed in biodiversity science. Beardsley and colleagues (2018) illustrated the need to educate both early-career and established biodiversity-oriented professionals in data and computational skills. Several initiatives are working to address this need, such as via iDigBio workshops (idigbio.org), the Carpentries (carpentries.org; Teal et al. 2015), and DataONE (dataone.org). Furthermore, the National Science Foundation (NSF)–funded project, Biodiversity Literacy in Undergraduate Education (BLUE, biodiversityliteracy.com) is uniting biodiversity and data scientists with education researchers to address the need for improved biodiversity data literacy specifically in undergraduate education. Per Barone and colleagues (2017), a survey of NSF-funded principal investigators agreed that they don't have the training or skills associated with using these complex data sets. Specialized training is needed to leverage biodiversity data resources for research in, and application to, complex environmental and resource management problems (Langen et al. 2014, Barone et al. 2017).

Addressing societal problems related to the environment and biodiversity requires an integrated research and training agenda that couples biodiversity and data science. This union provides critical skill sets for an empowered, twenty-first century science and engineering workforce to tackle our most complex challenges for sustainable use of natural resources, with applications to human health, agriculture, and climate resiliency. Leveraging resources and connecting biodiversity scientists and engineers via integrated training can hasten the development of a workforce to enable the next generation of biodiversity science. But this training must begin early. Across the science curriculum, biological informatics and data science are being prioritized in order to adequately train the twenty-first century workforce (see National Academies of Sciences, Engineering, and Medicine 2018). Data science education is directly addressed in K–12 (see Common Core Mathematics Standards, e.g., National Governors Association Center for Best Practices & Council of Chief State School Officers 2010, and Next Generation Science Standards, e.g., NGSS Lead States 2013) as well as in undergraduate education with the newly created

Four-Dimensional Ecology Education Framework from the Ecological Society of America and partners that incorporates data science into undergraduate ecology education, and as part of the recently developed BioSkills Guide developed to address recommendations of the AAAS's (2011) *Vision and Change in Undergraduate Biology Education*.

In undergraduate STEM fields, we have an opportunity to address workforce capacity by specifically using biodiversity content as a foundation for informatics and data science training priorities. Not only do students have insufficient data skills, but this deficit compounds an existing lack of understanding of the importance of biodiversity. Fortunately, curricula enriched with biodiversity data can address both of these challenges. To better take advantage, we must identify core content and skills that will empower undergraduates with a fundamental understanding of biodiversity science, as well as the skills necessary to discover and use resources for biodiversity research. The current concern over our changing planet means that all students—not only those in STEM fields—deserve an understanding of biodiversity-relevant concepts and processes in order to enable personal decision-making and participation in civic and cultural affairs. Students composing the twenty-first century workforce must be data literate, and providing data literacy in the context of biodiversity literacy will help our society effectively respond to novel diseases, increase sustainable agriculture to feed a growing global population, improve resiliency and adaptation to climate change, maintain a healthy planet, and so on. Natural history collections are a natural partner if we aim to use biodiversity content as a foundation for data literacy; they provide a data-rich resource that many different workforce sectors can identify as relevant to incorporating biodiversity services into their activities. Moreover, as was discussed elsewhere (e.g., Cook et al. 2014, Lacey et al. 2017, Monfils et al. 2017), these collections provide both tangible, place-based specimens and digital records for use in evidence-based science education and inquiry.

The twenty-first century workforce faces a new professional landscape that includes enhanced technology and connectivity, an increasing diversity of resources, and an explosion of available data. Educational approaches that include teaching about biodiversity data and content have a broad advantage both in training new biodiversity data users and in providing skills that are transferable to other disciplines. BLUE, as was described above, aims to improve undergraduate biology training by developing competencies, materials, and strategies for incorporating biodiversity data into the curriculum, and by facilitating broad-scale adoption of biodiversity data literacy goals. BLUE focuses on the introductory undergraduate level, providing broad impact through the training of educators, future scientists, and a data-literate workforce.

BLUE seeks to accomplish these goals by cultivating a diverse and inclusive network of biodiversity researchers, data scientists, and biology educators focused on undergraduate data-centric biodiversity education; building community consensus on core biodiversity data literacy competencies; developing strategies and exemplar materials to guide the integration of biodiversity data literacy competencies into introductory undergraduate biology curricula; and engaging a broader community of undergraduate educators in biodiversity data literacy efforts.

There is a gap between the level of training that most STEM undergraduates receive and the level of data literacy required in the twenty-first century workforce. Biodiversity-related issues can serve as a tangible and robust entry into data literacy. With this in mind, BLUE seeks to address the gap by defining core biodiversity data literacy competencies that students should develop during their undergraduate study and by identifying strategies for integrating these into an already crowded curriculum. Natural history collections and their associated data are a cornerstone of biodiversity science. They can be used to support authentic and inquiry-driven lessons in undergraduate biology (Cook et al. 2014, Lacey et al. 2017, Monfils et al. 2017), and are equally well suited to integration with other key data resources, e.g., National Ecological Observatory Network (NEON, neonscience.org).

In order to efficiently bridge the gap between the *research on*, and *practice of*, biodiversity data literacy, BLUE is incorporating best practices developed in the field of implementation science (Fixsen et al. 2017). This includes bringing together relevant community stakeholders, including biodiversity scientists, data scientists, educators, and education researchers to assess learning needs prior to the implementation of modules into curricula. For example, BLUE uses assessment data from listening sessions, one-on-one interviews, and surveys to help identify barriers to implementation. This then informs future module development and support for instructors, including resources that facilitate educators and students at generating their own hypotheses about a variety of evolutionary, ecological, and environmental questions and testing their hypotheses using open biodiversity data from freely available online sources. The BLUE strategy also involves growing and supporting a community of professionals interested in advancing biodiversity data literacy through hosting workshops, partnering with other science literacy initiatives, creating modules for undergraduate curricula, and spearheading the use of biodiversity collections data in education.

As tools to study Earth systems and global change biology become more sophisticated and readily available, the use of data-rich examples can truly transform the learning experience. Open data and open science beg for similar innovations in education, and BLUE seeks to energize and enable broad-scale adoption of biodiversity data and content for this purpose. BLUE is creating a community of scholarship across the fields of data science, biodiversity research, and science education to inclusively and synergistically optimize efforts, centralize resources, and facilitate the implementation of novel, community-vetted materials.

Student-centered approaches and relevant real-world examples, such as those offered by natural history collections and biodiversity-related topics, likewise provide opportunities to reach students who have not previously been effectively engaged in science courses or careers. The subsequent data-enabled community of STEM researchers and educators will be capable of shaping the biodiversity science of the future, as well as meeting the societal needs of a changing planet.

BLUE is an inclusive, open community, and we welcome your involvement. Find out more on our website, biodiversityliteracy.com.

Acknowledgment

The material presented here is based, in part, on a workshop that was funded by the National Science Foundation, grant no. DBI-730526.

References cited

[AAAS] American Association for the Advancement of Science. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action. AAAS. http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf.

Barone L, Williams J, Micklos D. 2017. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLOS Computational Biology 13 (art. e1005755).

Beardsley TM, Gropp RE, Verdier JM. 2018. Addressing biological informatics workforce needs: A report from the AIBS Council. BioScience 68: 847–853. https://doi.org/10.1093/biosci/biy116.

Biodiversity Collections Network. 2019. Extending U.S. Biodiversity Collections to Promote Research and Education. American Institute of Biological Sciences.

Cook JA, et al. 2014. Natural history collection as emerging resources for innovative education. BioScience 64: 725–734.

Fixsen DL, Blase KA, Fixsen AAM. 2017. Scaling effective innovations. Criminology and Public Policy 16: 487–499.

[IPBES] Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. 2019. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Brondizio ES, Settele J, Díaz S, Ngo HT, eds. IPBES Secretariat.

Lacey EA, Hammond TT, Walsh RE, Bell KC, Edwards SV, Ellwood ER, Guralnick R, Ickert-Bond SM, Mast AR, McCormack JE, Monfils AK. 2017. Climate change, collections and the classroom: Using big data to tackle big problems. Evolution: Education and Outreach 10: 2. doi:10.1186/s12052-017-0065-3.

Langen TA, Mourad T, Grant BW, Gram WK, Abraham BJ, Fernandez DS, Carroll M, Nuding A, Hampton SE. 2014. Opportunities and challenges in using large public data sets in the undergraduate ecology classroom. Frontiers in Ecology and the Environment 12: 362–363.

Marshall CR, et al. 2018. Quantifying dark data in museum fossil collections as palaeontology undergoes a second digital revolution. Biology Letters 14: 20180431.

McLean BS, Bell KC, Dunnum JL, Abrahamson B, Colella JP, Deardorff ER, Weber JA, Jones AK, Salazar-Miralles F, Cook JA. 2015. Natural history collections-based research: Progress, promise, and best practices. Journal of Mammalogy 97: 287–297.

Meineke EK, Davies TJ, Daru BH, Davis CC. 2018. Biological collections for understanding biodiversity in the Anthropocene. Philosophical Transactions of the Royal Society B 374 (art. 2017.0386).

Monfils AK, Powers KE, Marshall CJ, Martine CT, Smith JF, Alan Prather L. 2017. Natural history collections: Teaching about biodiversity across time, space, and digital platforms. Southeastern Naturalist 16: 47–57.

NGSS Lead States. 2013. Next Generation Science Standards: For States, By States. The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2018. Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report. National Academies Press.

National Governors Association Center for Best Practices and Council of Chief State School Officers. 2010. Common Core State Standards for Mathematics.

Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, Pawlik A. 2015. Data carpentry: Workshops to increase data literacy for researchers. International Journal of Digital Curation 10: 135–143.

Webster MS. 2017. The extended specimen. Pages 1–9 in Webster MS, ed. The Extended Specimen: Emerging Frontiers in Collections-Based Ornithological Research. Studies in Avian Biology no. 50. CRC Press.

Elizabeth R. Ellwood (ellwoodlibby@gmail.com) is affiliated with the Natural History Museum of Los Angeles County, La Brea Tar Pits, in Los Angeles, California. Jocelyn Anne Sessa is affiliated with the Department of Biodiversity, Earth, and Environmental Science at the Academy of Natural Sciences of Drexel University, in Philadelphia, Pennsylvania. Joel K. Abraham is affiliated with the Department of Biological Science at California State University, Fullerton. Amber E Budden is affiliated with the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara. Natalie Douglas is affiliated with the Department of Communication Sciences and Disorders at Central Michigan University, in Mt. Pleasant. Debra Linton is affiliated with the Department of Biology at Central Michigan University, in Mt. Pleasant. Robert Guralnic, Molly Phillips, and Pamela S. Soltis are affiliated with the, Florida Museum of Natural History at the University of Florida, in Gainesville. Erica Krimmel is affiliated with the Institute for Digital Information and Scientific Communication at Florida State University, in Tallahassee. Tom Langen is affiliated with the Department of Biology at Clarkson University, in Potsdam, New York. Marie Studer is a biodiversity education consultant in Stow, Massachusetts. Lisa D. White is affiliated with the Museum of Paleontology at the University of California, Berkeley. Jason Williams is affiliated with the Cold Spring Harbor Laboratory's DNA Learning Center, in Cold Spring Harbor, New York. Anna K. Monfils is affiliated with the Department of Biology at Central Michigan University, in Mt. Pleasant. ERE and AKM contributed equally to the production of the manuscript.

doi:10.1093/biosci/biz147