# Collections Education: The Extended Specimen and Data Acumen

ANNA K. MONFILS⬤, ERICA R. KRIMMEL, DEBRA L. LINTON, TRAVIS D. MARSICO, ASHLEY B. MORRIS, AND BRAD R. RUHFEL

*Biodiversity scientists must be fluent across disciplines; they must possess the quantitative, computational, and data skills necessary for working with large, complex data sets, and they must have foundational skills and content knowledge from ecology, evolution, taxonomy, and systematics. To effectively train the emerging workforce, we must teach science as we conduct science and embrace emerging concepts of data acumen alongside the knowledge, tools, and techniques foundational to organismal biology. We present an open education resource that updates the traditional plant collection exercise to incorporate best practices in twenty-first century collecting and to contextualize the activities that build data acumen. Students exposed to this resource gained skills and content knowledge in plant taxonomy and systematics, as well as a nuanced understanding of collections-based data resources. We discuss the importance of the extended specimen in fostering scientific discovery and reinforcing foundational concepts in biodiversity science, taxonomy, and systematics.*

*Keywords: biodiversity science, community science, herbaria, natural history collections, undergraduate education*

**I**n biodiversity science, a multidisciplinary field focused on understanding life on Earth, we work across disciplines, including but not limited to environmental science, evolutionary biology, systematics, ecology, and paleontology. We increasingly employ quantitative reasoning, computational skills, and data acumen (the ability to evaluate and use data effectively) to conduct research using large and complex data sets bridging time, space, and scale (Hampton et al. 2017, NASEM 2018, Ellwood et al. 2020). This represents an expansion of rather than a shift in the skills and concepts we expect current biodiversity scientists to possess and future scientists to gain. As educators training the twenty-first century biodiversity workforce, we can rise to the occasion and reimagine our courses to teach new skills and content in support of data acumen while simultaneously retaining foundational skills and content from organismal biology (Wen et al. 2015, Bik 2017, Raven and Miller 2020). This can be challenging in a crowded undergraduate biology curriculum that is increasingly becoming depauperate of organismal courses (Middendorf and Pohlad 2014, Crisci et al. 2020). As scientists, researchers, and educators, we must be intentional as we integrate the skills, knowledge, and tools to harness the data revolution while fostering the core taxonomic, systematics, and organismal biological knowledge necessary to address emerging questions in biodiversity science.

Biodiversity science has seen the volume and variety of data available for analyses increase at a meteoric pace (Hampton et al. 2013 and 2017, Jarić et al. 2020). Globally, there have been massive advances in digitizing specimen-based data from biological collections, including projects such as the US National Science Foundation's Advancing Digitization of Biodiversity Collections program, the Global Biodiversity Information Facility, the European Union's Distributed System of Scientific Collections, and the Atlas of Living Australia. Observations of organisms contributed by community scientists are growing via portals such as iNaturalist and the US National Phenology Network. Organismal data are in turn centralized by organizations such as the International Union for Conservation of Nature and NatureServe, which package them for use by policy and land management practitioners. In the United States, associated data on climate, weather, land use, soil, and water are available online from federally funded departments including the National Oceanic and Atmospheric Administration and the US Geological Survey. These broadscale data sources are complemented by targeted databases such as the Global Soil Dataset for Earth System Modeling. Projects such as the US National Ecological Observatory Network work to connect existing environmental data layers to field-collected specimens and long-term observational species occurrence data. Biodiversity science is faced with the exciting prospect of pursuing integrative research spanning biotic and abiotic data sets at unprecedented resolution across multiple scales, making possible the synthetic analyses that can address pressing

Advance Access publication 27 October 2021

environmental and ecological problems (LaDeau et al. 2017, Farley et al. 2018, Henkhaus et al. 2020, Wüest et al. 2020).

Biological collections serve as critical research infrastructure in the data-rich field of biodiversity science and have played a key role in transforming twenty-first century biology (Wen et al. 2015, Page et al. 2015, Lendemer et al. 2020, Monfils et al. 2020, NASEM 2020). The specimens curated in collections provide a source of data that is unparalleled in temporal, geographic, and taxonomic complexity and is unique in its ability to allow researchers to verify and expand data by returning to the physical specimens on which they are based (Suarez and Tsutsui 2004, Pyke and Ehrlich 2010, McLean et al. 2016, and the references therein). Physical specimens, preserved and curated in perpetuity, are the basis of scientific names and a foundation of scientific discovery both within and beyond biology (Funk 2018, NASEM 2020). Specimens provide a valuable source of biodiversity data, including molecular, biochemical, and species associations from the physical vouchers, and spatiotemporal, cultural, and ecological information from the collection records. In addition, specimens can be digitally linked to each other and to an increasing number of other data resources (Schindel and Cook 2018). The term *extended specimen* was coined to refer to the specimen, associated specimen data, and interconnected network of data resources, which, together, have incredible potential to enhance integrative and data-driven research (Webster 2017, Hardisty et al. 2019, Lendemer et al. 2020). Specimens can function as bridges between disciplines and databases (genetic, ecological, conservation, environmental, paleontological; Wen et al. 2015). In the last decade, digitized and online specimen-based data have greatly increased our capacity to investigate large-scale issues of critical importance in the twenty-first century—for example, climate change and its impacts, zoonotic disease transmission, sustainable resource management, impacts of invasive species, and biodiversity loss (Ball-Damerow et al. 2019, Heberling et al. 2019, Cook et al. 2020, Thompson et al. 2021).

The current and next generations of biodiversity scientists need extended training in order to be prepared to contribute to and conduct research using extended specimen data (Barone et al. 2017). Collection of new specimens remains critical to biodiversity science, and new twenty-first century collecting practices have been adapted and must be employed to support the born-digital and linked data that define the extended specimen (Miller et al. 2020). Twenty-first century collecting involves complex specimen collection strategies, preparation techniques, and data management planning (Schindel and Cook 2018). To stay current, biodiversity science workforce training can incorporate more computational skills and technologies such as those developing in the domains of data science and biological informatics (including bioinformatics; Hampton et al. 2017, Beardsley et al. 2018, Robeva et al. 2020). At the same time, foundational skills related to understanding organismal biology and taxonomy remain essential (Bik 2017). The

need for biodiversity scientists grows, but undergraduate education globally is facing a disconcerting decrease in taxon-specific coursework, field biology, and taxonomy, which all have important content needed for the next generation of biologists to identify, describe, name, and archive biodiversity (Britz et al. 2020). While reinforcing the importance of data acumen in biodiversity science, we cannot allow the education pendulum to swing toward emerging technologies and skills at the expense of foundational skills from field biology, taxonomy, and systematics. Now, more than ever, we need to be creative and intentional in providing professional development opportunities for researchers and educators by employing best practices in undergraduate biology education, so that we can facilitate a biodiversity-literate *and* data-literate workforce.

Across all the STEM disciplines, there is a pressing need for researchers to gain data skills (Barone et al. 2017) and educators to incorporate data acumen into their undergraduate instruction (AAAS 2011, NITRD 2016, NASEM 2018). Physical natural history specimens combined with extended data can provide an engaging and concrete resource that is well suited to authentic student-centered learning and translatable data acumen (Cook et al. 2014, Powers et al. 2014, Monfils et al. 2016, Lacey et. al 2017, Ellwood et al. 2020). Regardless of the discipline, the extended specimen concept and requisite twenty-first century specimen collecting practices provide an ideal teaching tool for introducing STEM educators and students to data sources and skills. When teaching with the extended specimen concept, students collecting their own specimens engage in the full data lifecycle, including data and metadata collection, preparation and archiving of specimens, and submission of data to an appropriate aggregator or repository. Once submitted, the students can track the usage of their data and also capitalize on complementary data sets to ask integrative research questions. As the students work with specimens and their data, they are exposed to important concepts in data acumen, such as provenance, metadata, standards, and validation. The students' close association as collectors of the specimens and stewards of the specimen data can make these data concepts more accessible and meaningful (Ellwood et al. 2020).

In the present article, we provide an example of how a team of scientists and educators was able to use best practices in undergraduate education and in twenty-first century specimen collecting to incorporate evolving workforce skills (e.g., data acumen) into our courses without losing attention to foundational skills and concepts in biodiversity science. Building off of our experience and training in specimen digitization, we redesigned the classic plant collecting exercise in which students collect, press, and identify "unknown" plant species in order to gain foundational skills in species identification and taxonomy. Our modified module reframed this exercise on the basis of recommendations for twenty-first century specimen collecting (e.g., from Miller et al. 2020) and additionally

---

**Box 1. Learning objectives of the "Connecting students to citizen science and curated collections" module.**

Students completing this assignment will be able to (1) maintain a professional specimen collection notebook, (2) collect plant specimens from the field using proper techniques and including adequate material for identification, (3) prepare and deposit research quality herbarium specimens, (4) deposit species and occurrence data into national or international databases, (5) identify unknown plant specimens using multiple forms of reliable evidence, (6) explain the importance of herbaria in plant biology research, (7) discuss the value of large data sets for investigating spatially or temporally large phenomena (invasive species, rare species, climate change, etc.), and (8) evaluate the importance of citizen scientists to large data sets.

Since developing and assessing this module, there has been a collective shift moving away from the term *citizen science* to more inclusive terminology. When discussing the field of research within this manuscript we have adopted the term community science but have retained the term *citizen science* when referencing defined learning outcomes, assessment, and published OER resources.

---

incorporated data acumen into our learning objectives. Working as a community of practice (Wenger 2011, AAAS 2011), we identified and assessed our new and foundational learning outcomes, developed and implemented educational materials, and shared our educational products as an open education resource (OER; Biswas-Diener and Jhangiani 2017). Provided below is a description of our module development, assessment, and learning outcomes as a proof of concept in support of a broader discussion on merging advances in biological collections and in education in support of an integrated biodiversity science discipline and data-savvy workforce.

## Example module

The "Connecting students to citizen science and curated collections" module (www.collectionseducation.org) originated at a North American Network of Small Herbarium digitization workshop hosted by Integrated Digitized Biocollections (iDigBio) at the Botany 2014 annual conference in Boise, Idaho. During module development we identified two primary goals for the project: to educate students in evolving best practices in specimen collection and curation and to reduce the creation of future specimen processing backlogs resulting from analog student collections that require transcription to create digital records. We recognized that twenty-first century specimen collecting practices take advantage of born-digital data (e.g., generating label data in a digital format, or photographing specimens *in situ*; see Hackett et al. 2019, Powell et al. 2019) and that specimen collecting could be used as a platform for educating students in emerging data skills and concepts related to open science, digital data resources, and community science (note that this term is a more inclusive replacement for *citizen science*, recognizing the role of all participants regardless of citizenship; Audubon Society 2021). We based the module on a collection curation workflow that leverages born-digital data and emphasizes current best practices both in botanical collecting and in specimen data management; these best practices align with the digital data literacy workforce training recommendations for the extended specimen (Lendemer et al. 2020, NASEM 2020). Our module employed the iNaturalist platform

(www.iNaturalist.org) for recording data and images at the site of the collecting event, as has become popular with professional botanists (e.g., Heberling and Isaac 2018). From inception, we sought to engage in the scholarship of teaching with the goal of publishing our assessed and vetted educational materials. We were deliberate in developing online educational materials in the spirit of open science and FAIR practices (for *findable*, *accessible*, *interoperable*, and *reusable*; Wilkinson et al. 2016, Garcia et al. 2020).

**Module and assessment design.** We used standard practice in backward design to create the module (Wiggins and McTighe 2005; see Linton et al. 2020 for a detailed workflow), beginning by identifying our desired results in the form of learning objectives that address foundational skills in species identification and twenty-first century specimen collecting (see box 1). On the basis of these learning objectives, we designed a student assignment with pre- and postmodule assessment to serve as acceptable evidence for evaluation (see supplement 1). The assignment included an instructor-defined number of plant collections, each of which included an archival-quality herbarium specimen, an iNaturalist observation entry, and detailed specimen-specific plant identification notes (see supplement 2). The pre- and postmodule assessment included seven free response questions about the students' knowledge of key concepts and four Likert-scale questions about the students' opinions (see supplement 3). The premodule assessment also included general questions about the students' academic status, prior course work, and plant collecting experience. In the postmodule assessment, additional Likert-scale questions asked the students to self-report how well the module prepared them to perform the skills identified in the learning objectives. The appropriate IRB approval was received for each institution at which the assessment was implemented (Arkansas State University IRBNet646782, Central Michigan University IRB-661959-1, Eastern Kentucky University IRB-15-110, Middle Tennessee State University IRB-15-032).

Assessments were evaluated after the conclusion of the 2-year initial implementation period. Coding rubrics for evaluating the free response questions were developed by consensus among five of the project collaborators with

expertise in plant biology and information science, on the basis of key concepts that all agreed would be included in an "expert" answer, as well as specific misconceptions (see supplement 4). Hardcopy student responses were deidentified, transcribed, and assigned to experts not associated with the student's institution. Two experts coded each question according to the relevant rubric. If the ratings from the two experts did not match, the response was sent to a third expert for rating, and the majority rating for each item was recorded as the consensus coding.

For each free response question, the total number of correct concepts per student was compared before and after the module was implemented using the Wilcoxon signed-rank test (Wilcoxon 1992). For specific individual responses within a question, the percentage of the students who included a specific correct response was compared pre- and postmodule using McNemar's test of correlated proportions (McNemar 1947). When evaluating Likert scale questions, the Likert responses were converted to numerical values (e.g., 5, *necessary*; 4, *very important*; 3, *important*; 2, *somewhat important*; 1, *not at all important*), and significant shifts in the student responses before and after the module were compared using the Wilcoxon signed-rank test.

**Module implementation and outcomes.** The module is hosted on a public website, www.collectionseducation.org, and is published as an OER via Quantitative Undergraduate Biology Education and Synthesis at https://qubeshub.org/publications/1103 (Monfils et al. 2019), where it is presented in a set of educational resources available through the Biodiversity Literacy in Undergraduate Education: BLUE Data Network (www.biodiversityliteracy.com), a National Science Foundation Research Coordination Network in Undergraduate Biology Education. The module website was launched in September of 2014 and, by 2021, had been visited over 80,000 times. We published the module on QUBESHub.org in March of 2019, and by August of 2021, it had been accessed over 1100 times.

The module was implemented and assessed in 10 courses: plant systematics (four classes), dendrology (three classes), aquatic and wetland plants (two classes), and flowering plants (one class) at four participating universities, ranging from large master's level colleges and universities (M1) to doctoral universities with high levels of research (R2; Carnegie classifications, Indiana University 2017, https://carnegieclassifications.iu.edu; Arkansas State University was R3 at the time of module implementation; Central Michigan University, R2; Eastern Kentucky University, M1; and Middle Tennessee State University, M1) between Fall 2014 and Spring 2016. A total of 148 students signed the appropriate IRB consent form, completed all assessments and were included in the data analyses. This sample population was 82% undergraduate and 14% graduate student (4% other). For the students enrolled in the courses, 51% had taken a prior plant identification course, and 28% indicated they had prior course experience with a plant collection exercise.

The results from the pre- and postmodule assessments indicated significant gains in self-efficacy and content knowledge (figures 1–6). The educators employing the module commented on the high-quality specimens, iNaturalist observations (figure 1a), and specimen labels (figure 1b) resulting from the student engagement in the plant collecting and identification exercise. The educators also noted a decrease in specimen backlog and digital curation needs for their respective herbaria. In the present article, we discuss key results as they relate to our module learning objectives.

The students maintained specimen collection notebooks, used proper field collection techniques for identifying and archiving specimens, and were able to prepare and deposit quality herbarium specimens into a physical collection, as well as deposit associated specimen data into national or international databases. In our postmodule assessment, the students felt well prepared, very well prepared, or totally prepared to use foundational and emerging plant collecting skills including maintaining a field notebook (89%), collecting specimens in the field (94%), and depositing specimens (89%) and digital data (92%) into national and international data repositories (figure 2a–2d). In addition, the instructors implementing this module across the four universities and 10 courses indicated that the students' specimens and labels were of higher quality (i.e., they contained more explicit information and were formatted better) than the labels of the students participating in plant collection and identification exercises not following this module. The students completing this module showed confidence in their collection skills and educators noted an overall increase in quality of specimens and associated data compared with classes not using this module.

The students recognized plant identification resources, gained confidence in plant identification skills, and self-identified as well prepared to apply these skills. In our postmodule assessment, over 90% of our students indicated they felt well prepared, very well prepared, or totally prepared to identify unknown plant specimens using multiple forms of reliable evidence (figure 2e). When the students were asked to list the steps in plant identification and identify resources for verification, they identified significantly more verification resources postmodule compared with premodule (4.0 versus 2.9; $p < .0001$). The students' responses included specific resources, where we saw a significant increase in responses identifying taxonomic keys (61% to 82%; $p < .0001$), herbarium specimen comparison (9% to 22%; $p = .001$), image comparison (41% to 56%; $p = .0002$), and expert determination (14% to 33%; $p < .0001$; figure 3). The instructors noted that the students showed increased engagement and comfort with plants and plant identification. The students' ability to identify appropriate resources for plant identification, as well as increased confidence in their ability to correctly identify plant species, makes plants more visible and plant knowledge more accessible to them

**(a) iNaturalist observation**

**(b) herbarium specimen label**



*Figure 1. Example iNaturalist observation (a) and subsequent herbarium specimen label (b) created as part of this module, illustrating what data are recorded where, including (1) photographs of the specimen in situ, (2) detailed notes about the collecting event, (3) taxonomic identifications by the collector and subsequent identifiers, (4) geographic coordinates of the collecting site, and (5) metadata about associated species, collector name and number, elevation, habitat, and precise collecting locality that will be printed on the herbarium label. This record can be viewed in full at www.inaturalist.org/observations/8027978.*

and addresses an overarching concern regarding plant awareness disparity (i.e., the lack of knowledge, interest, and perception of plants in comparison to animals; Wandersee and Schussler 1999, Parsley 2020).

The students were able to explain the importance of herbaria in plant biology research. This included an increased and nuanced understanding of what an herbarium is, who can use herbaria and associated data, and how herbarium collections and collections-based data can be used in research. Postmodule assessment results showed 91% of the students from our study felt well prepared, very well prepared, or totally prepared to explain the importance of herbaria in plant biology research (figure 2f). After the module was implemented, we saw a significant increase in the students correctly stating that an herbarium is a collection of preserved plants (62% to 98%; $p < .0001$) and a significant decrease in the misconception that herbaria contain living plants (22% to less than 1%; $p < .0001$). There was also a significant increase in the students stating that

herbaria contain valuable data (7% to 18%; $p = .002$) and can be used for research (17% to 26%; $p = .027$).

The students indicated before and after the module that it was important, very important, or necessary for biologists (99% to 99%; $p = .2119$), scientists (90% to 89%; $p = .3264$), and educators (92% to 90%; $p = .2546$) to have a strong understanding of the potential uses of biological collections (figure 4a). These values were not significantly different between the pre- and postmodule assessment. We did, however, see a significant positive shift between before the module and afterward in the students noting that all citizens need to understand potential uses of biological collections (54% to 58%; $p = .015$; figure 4a).

When the students were asked who uses herbaria, they identified significantly more types of professionals who might work with the collections on the postmodule assessment (2.6 to 3.3; $p < .0001$). The majority of the students both before and after the module identified plant scientists or botanists (74% to 81%; $p = .055$) and biologists

**Student responses to the question "Please rate how effectively the specimen and data collection project has prepared you to..."**
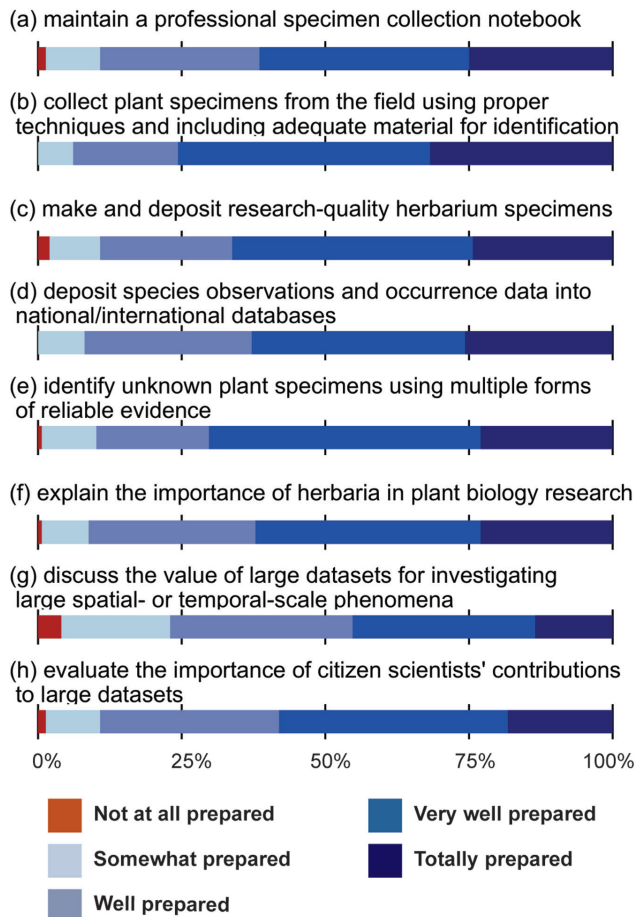
(a) maintain a professional specimen collection notebook

(b) collect plant specimens from the field using proper techniques and including adequate material for identification

(c) make and deposit research-quality herbarium specimens

(d) deposit species observations and occurrence data into national/international databases

(e) identify unknown plant specimens using multiple forms of reliable evidence

(f) explain the importance of herbaria in plant biology research

(g) discuss the value of large datasets for investigating large spatial- or temporal-scale phenomena

(h) evaluate the importance of citizen scientists' contributions to large datasets

0%          25%          50%          75%          100%

- ■ Not at all prepared
- ■ Somewhat prepared
- ■ Well prepared
- ■ Very well prepared
- ■ Totally prepared

*Figure 2. Student self-assessment of learning gains after the module.*

**Student responses to the question "Explain the steps you would take to identify an unknown plant specimen. List all the resources you know of that can help you verify your identification."**
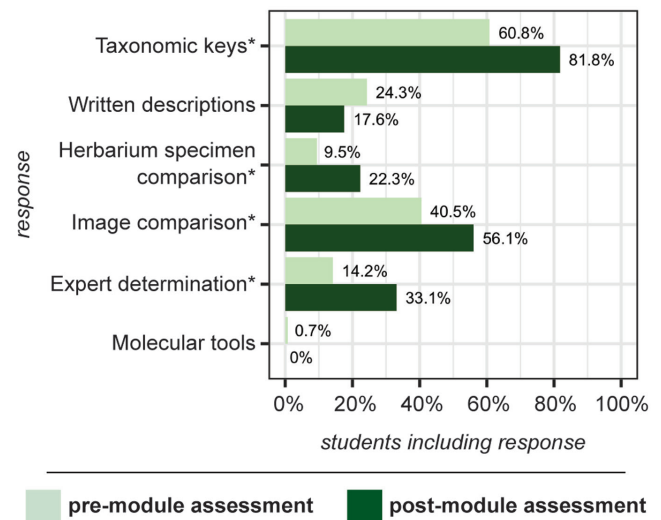
*Taxonomic keys\** — 60.8% / 81.8%
*Written descriptions* — 24.3% / 17.6%
*Herbarium specimen comparison\** — 9.5% / 22.3%
*Image comparison\** — 40.5% / 56.1%
*Expert determination\** — 14.2% / 33.1%
*Molecular tools* — 0.7% / 0%

0%    20%    40%    60%    80%    100%
*students including response*

- ■ pre-module assessment
- ■ post-module assessment

*Figure 3. Pre- and postmodule assessment of student gains toward learning objective 5. Statistical significance is indicated by an asterisk (*$p \leq .05$).*

(45% to 71%; $p < .0001$) and environmental change (6% to 16%; $p = .002$; figure 4c). Although it was low in the initial premodule assessment, the postmodule assessment saw a significant decrease in the students' noting hypotheses related to disease (3% to 0%; $p = .031$) and medicine or ethnobotany (5% to 1%; $p = .016$; figure 4c). We suspect this response could change given the COVID-19 pandemic, and these themes would be worth highlighting in future courses. The students' breadth of response could be enhanced by exercises and educational materials that introduce the rich literature associated with use of collections.

For the students participating in our module, experience in and with herbaria and specimen collecting rectified the misconception that herbaria house living specimens and opened a new biodiversity tool and data resource the students could use when entering the workforce. The students' increases in understanding the relevance of collections to science and society is important when looking to new and future end users of biodiversity data.

The students gained a better comprehension of the value of aggregated biodiversity data sets for investigating spatially or temporally large phenomena. The students could describe differences between specimen- and observation-based natural history data, correctly identify an increased number and diversity of challenges associated with using large, aggregated data sets, and identify the importance of all citizens having knowledge of big data and their uses. Overall, 77% of the students from our study felt well prepared, very well prepared, or totally prepared to discuss the value of large data sets for investigating spatially or temporally large

(not specific to plant science or botany; 51% to 53%; $p = .38$) as professionals who might work with herbaria, but there was not a significant difference in these percentages before and after the module (figure 4b). We did see a significant increase from before to after the module in the students identifying researchers (not specific to biology; 11% to 18%; $p = .04$), educators (16% to 30%; $p = .001$), law enforcement professionals (1% to 4%; $p = .03$), and politicians or policy advocates (0% to 3%; $p = .03$) as potential users of herbaria (figure 4b).

The students were asked to list hypotheses a researcher could test using data collected from herbarium specimens. Not only did the students have a significant increase in the number of proposed hypotheses (2.1 to 2.7; $p < .0001$), but we also saw an increase in the number of different topics covered by the proposed hypotheses (1.7 to 2.2 different topics identified; $p < .0001$). The two categories of hypothesis topics with significant increases from before to after the module were related to species distributions

**(a) Student responses to the question "Please rate how important you think it is for the following groups to have a strong understanding of biological collections uses (e.g., uses of herbaria)."**



*1 = Not important, 2 = Somewhat important, 3 = Important, 4 = Very important, 5 = Necessary*
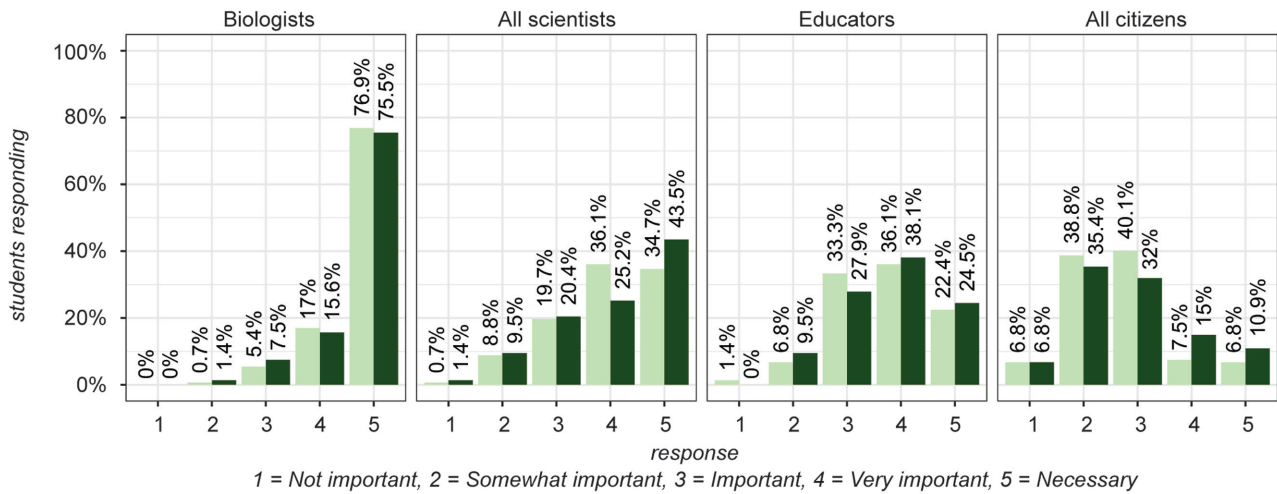
**(b) Student responses to the question "What type of career professionals use herbarium collections?"**
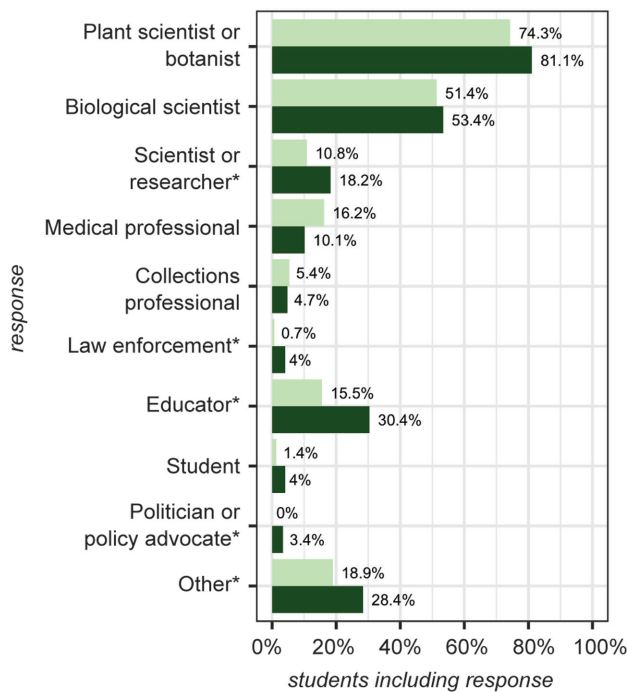


**(c) Student responses to the question "List three different hypotheses you could test (or three different questions you could answer) using data collected from herbarium specimens. "**
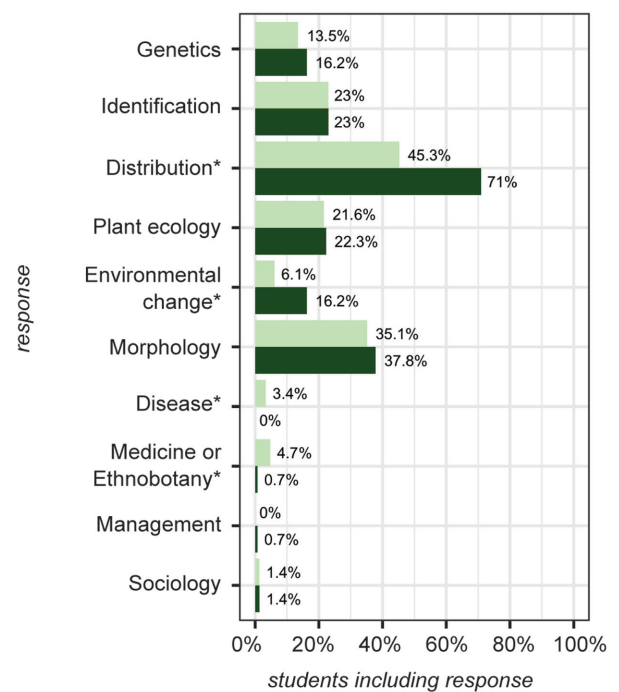


■ pre-module assessment   ■ post-module assessment

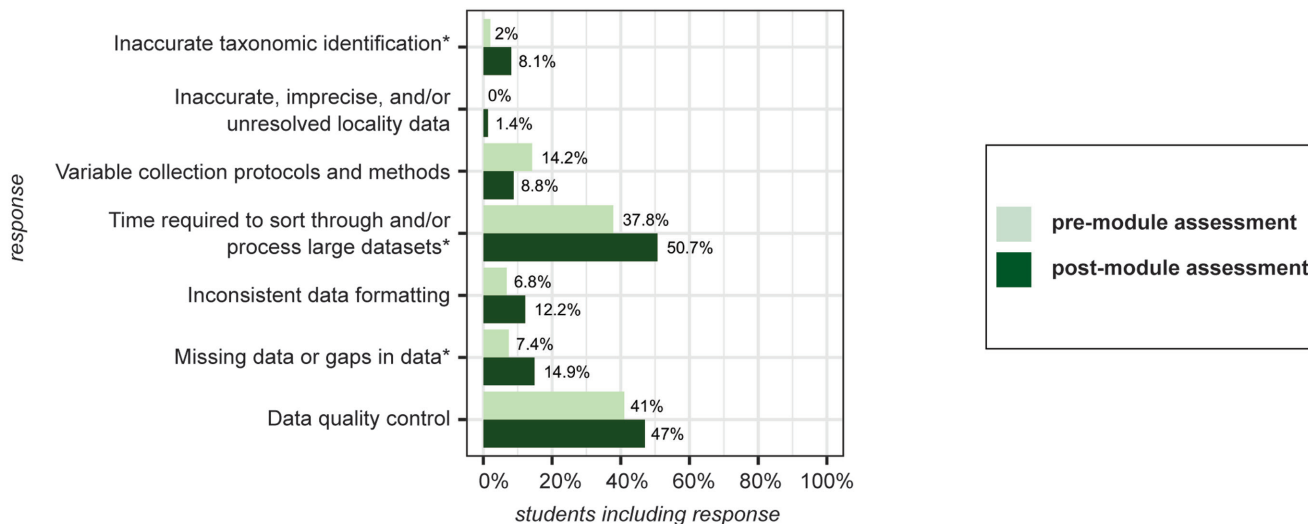*Figure 4. Pre- and postmodule assessment of student gains toward learning objective 6. Statistical significance is indicated by an asterisk (\*p ≤ .05).*

phenomena (figure 2g). We asked the students an open-ended question about differences between specimen-based and observation-based data and saw a significant increase in responses that noted either that specimen-based data were verifiable or that observation-based data were not verifiable

(35% to 51%; $p = .001$) indicating an understanding of biodiversity data types.

The student responses on the postmodule assessment showed an increase in their ability to identify challenges associated with large, aggregated data sets (figure 5a). There

**(a) Student responses to the question "What are some challenges associated with using large, aggregated datasets? Feel free to list challenges as assumptions that have to be made or limitations that must be considered before analyzing data for a research project."**



**(b) Student responses to the question "Please rate how important you think it is for the following groups to have a strong understanding of large datasets or 'big data' use and analysis."**
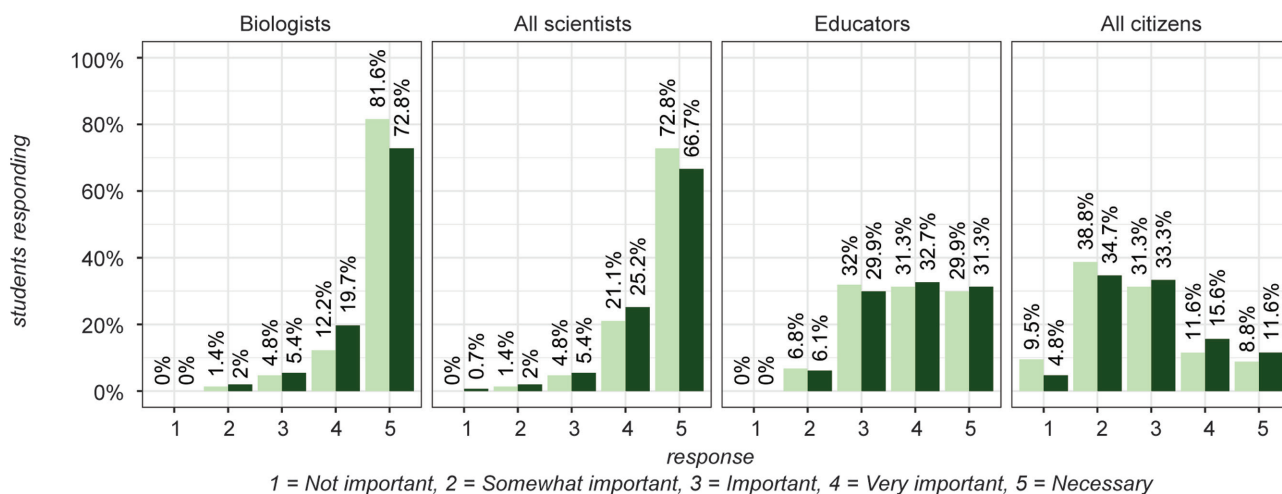


*Figure 5. Pre- and postmodule assessment of student gains toward learning objective 7. Statistical significance is indicated by an asterisk (\*p ≤ .05).*

was a significant increase in the percentage of the students noting inaccurate identification of specimens (2% to 8%; $p = .006$), time to process data (38% to 51%; $p = .006$), and missing data (7% to 15%; $p = .017$) as challenges. About half of the students also identified variable data quality as a challenge, but this value did not change significantly between the pre- and postmodule assessments (41% to 47%; $p = .13$).

The students were asked to rate how important it is for different groups of people to have a strong understanding of large data sets or big data use and analysis. Over 90% of the students rated knowledge about these types of data as

important, very important, or necessary to biologists (99% to 98%; $p = .0668$), scientists (99% to 97%; $p = .0548$), and educators (93% to 94%; $p = .2743$) both before and after the module (figure 5b). There was no significant change in these percentages before and after the module; however, we did see a significant shift toward greater importance of this knowledge for all citizens (52% to 61%; $p = .0075$; figure 5b).

In an effort to teach foundational knowledge regarding biodiversity informatics, it is critical that students have an understanding of data. Our module exposes undergraduate students to biodiversity data in a way that provides a

**(a) Student responses to the question "What are 'citizen scientists'?"**
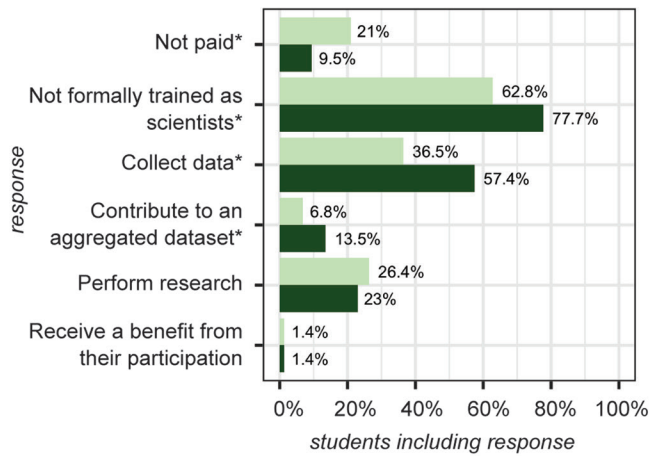


**(c) Student responses to the question "How important do you think citizen science activities will be for your future profession?"**
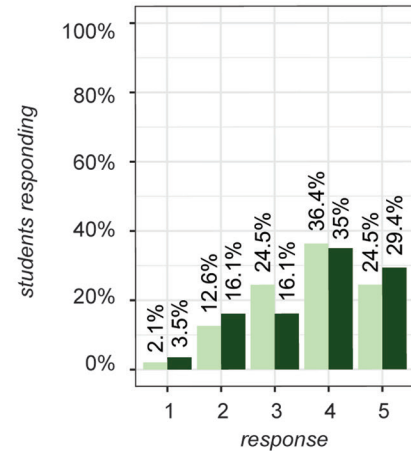


*1 = Not important, 2 = Somewhat important, 3 = Important, 4 = Very important, 5 = Necessary*

**(b) Student responses to the question "Please rate how important you think it is for the following groups to have a strong understanding of citizen science contributions."**
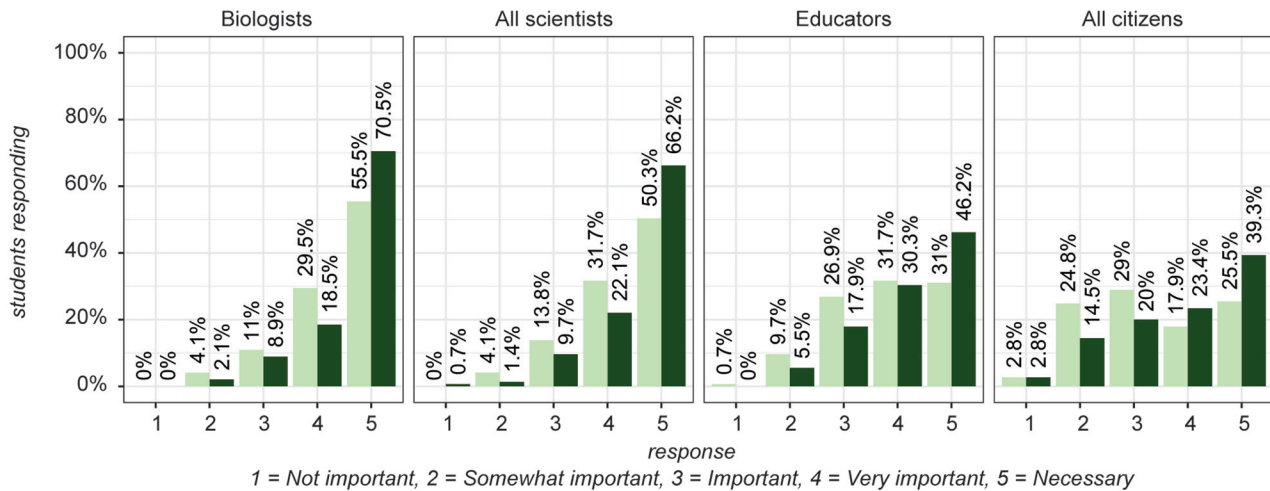


*1 = Not important, 2 = Somewhat important, 3 = Important, 4 = Very important, 5 = Necessary*

pre-module assessment    post-module assessment

*Figure 6. Pre- and postmodule assessment of student gains toward learning objective 8. Statistical significance is indicated by an asterisk (\*p ≤ .05).*

baseline understanding of and appreciation for the potential applications of these data, while also providing them with data acumen that is translatable and applicable to any modern career field.

The students increased their understanding of the role of citizen science in biodiversity science. (The students in the course and subsequent assessment used the term *citizen science*; we retain that terminology to align with the learning objectives and assessment. We use the more inclusive *community science* elsewhere in the article). In our study, 89% of the students felt well prepared, very well prepared, or totally prepared to evaluate the importance

of citizen scientists to large data sets (figure 2h). We asked the students to define what a citizen scientist is, and there was a significant increase in the percentage of the students who stated that citizen scientists are not formally trained as scientists (63% to 78%; *p* = .002) and a significant decrease in the percentage of the students noting citizen scientists are not paid (21% to 9%; *p* = .002; figure 6a). About a quarter of the students on the premodule assessment noted that citizen scientists perform research, and we saw no significant change on this after the module (26% to 23%; *p* = .026; figure 6a). However, there was a significant increase in the students' noting that citizen scientists

collect data (36% to 57%; $p < .0001$) and contribute to aggregated data sets (7% to 14%; $p = .02$; figure 6a).

The students indicated before and after the module that it was important, very important, or necessary for biologists (96% to 98%; $p = .0075$), all scientists (96% to 98%; $p = .0038$), educators (90% to 94%; $p = .0005$), and all citizens (74% to 83%; $p = .0001$) to have a strong understanding of the contributions of citizen science (figure 6b). These were significant positive shifts for all categories. The students were then asked to rate the importance of citizen science activities in their own future professions. Both before and after the module, more than three quarters of the students (82% to 77%; $p = .39$) rated citizen science activities as important, very important, or necessary for their future profession; there was no significant shift in the ratings scale (figure 6c).

The instructors noted the students' active engagement with iNaturalist and the community of iNaturalist users. The experience of participating in community science through our module provides a framework for students to understand this increasingly important source of biodiversity data and is something they can build on either in a future STEM-oriented career or as a biodiversity-literate member of society. There is some indication of a better understanding of community science and its contribution to biodiversity science. There is potential here for students to further explore the role of community science in biodiversity science and additional exercises would help reinforce students' understanding of community science contributions and potential.

The instructors saw a decrease in processing backlogs resulting from the student collections. The instructors reported that the module was successful from a specimen processing and digital curation perspective. Incorporating the student collections into the research herbarium or teaching collection was streamlined for the following reasons: All specimens had a digital record already present on iNaturalist, the presence of the digital record made specimen label formats detailed and uniform, and the specimens could be quickly sorted into research herbarium or teaching collection piles to be mounted and accessioned. Because the specimen processing was streamlined, subsequent students were able to take advantage of having recent collections to reference, both in the form of physical specimens and of digital iNaturalist observations with photos *in situ*. Even if a physical specimen was still awaiting accession, the iNaturalist record provided access to the data. This scenario exemplifies the open science tenet of "share early, share often" for the students.

We identified two primary goals for the project: to educate students in evolving best practices in specimen collection and curation and to reduce the creation of future specimen processing backlogs resulting from student collections. Across the individual courses, institutions, and 2-year timespan (2014–2016), our assessment data consistently showed that our "Connecting students to citizen science and curated collections" module increased student awareness of sources for plant identification and increased their confidence in their ability to correctly identify specimens. The students gained a better understanding of herbaria, and gained critical data acumen skills relative to data discovery, data types, data aggregation, and data usage. The students identified potential users of herbaria, and research questions that could be asked using aggregated herbarium specimen-based data. Specifically, the students could explain the difference among data types (specimen based, observation based, and community science derived) and could describe the challenges and assumptions of aggregated data sets. The students also increased in their belief that all citizens should have an understanding of large data sets and their analyses. By the end of the module, the students felt prepared for twenty-first century specimen collecting. Our assessment data provide evidence that the students completing this module learned foundational skills and content related to taxonomy, systematics, and field botany. The assessment data also indicate that the students gained emerging skills and exposure to concepts related to data acumen, enabling them to be better prepared for engaging with large, complex data sets involving the extended specimen.

## Conclusions

In the last decade, we have seen a transformation of undergraduate biology education. At the time this module was developed, in 2014, we had guidance from the recommendations of *Vision and Change in Undergraduate Biology Education: A Call to Action* that challenged undergraduate biology educators to teach biology as we do biology in the twenty-first century (AAAS 2011). Almost in parallel, we saw a massive explosion of open data and a prioritization of the importance of biodiversity data, specifically specimen-based data. As researchers and curators of regional collections, we were pushing ourselves to learn digitization processes and new computational and informatics skills related to creating, managing, and accessing digital data. We made a conscious decision to leverage the emerging data skills and technologies we were learning through digitizing our respective regional collections in our approach to teaching plant taxonomy, systematics, and field biology. We modeled the community of practice and collaboration we were using in the national digitization effort with how we built and implemented our educational module to effectively meet our research needs and our teaching efforts. Our approach kept us current and forward thinking. By merging advances in informatics and biodiversity databases, the breadth of physical specimen data, our collective expertise in plant biology, and changes in twenty-first century science with the teaching and learning recommendations of *Vision and Change* (AAAS 2011), we leveraged the scholarship of teaching and of science to create real educational experiences where the students learn science by practicing science at the cusp of science discovery.

As we continue to be challenged in how we integrate new skills and content into the curriculum, our example module's

plant collection and identification exercise demonstrates how we can integrate advances in the practice of science into curricula while continuing to engage the students in foundational science practices (e.g., species literacy, taxonomy, plant identification). The OER resources can help educators acquiring new digital curation skills and can assist current professionals entering digital biodiversity science research. By employing an open education framework, we encourage the continued adaptation and modification of this module to include important, emergent themes in biodiversity science and data acumen—for example, global biodiversity knowledge (Hobern et al. 2019), decolonization of collections (Turner 2015), traditional ecological knowledge and conservation (Biró et al. 2014), and the CARE Principles for Indigenous Data Governance (Carroll et al. 2020).

We used the same core foundational science practices supporting the extended specimen (open science, open data, twenty-first century collections, and interdisciplinary collaboration; Lendemer et al. 2020) to build a community of practice for education, share resources openly, build and reinforce digital data skills and content knowledge, and enhance our students' skills related to taxonomy and systematics. In turn, our team of educators was able to enhance the quantity and quality of digitized data we mobilized from our herbaria, our research expertise, and the research contributions in our respective labs. Our students were able to engage with the full data lifecycle and recognize opportunities for interdisciplinary thinking in biodiversity science—for example, between aggregated data resources or with new audiences such as the community scientists using iNaturalist. The extended specimen offers an opportunity to develop extended resources for educators, extend skills that our students can gain in core biodiversity and data science, extend data resources and collections, and extend research capacity and questions we can address.

Students today must leave their undergraduate education with the ability to see the potential of open, aggregated databases; understand the tools to access, analyze, and infer information from data; and engage with a diverse, interdisciplinary community to make strides in addressing the biodiversity crises of the Anthropocene. As scientists and educators, we must embrace the changing landscape of biodiversity science and leverage the foundational skills that collections have fostered for centuries to help engage, inspire, and build the next generation of scientists. Our "Connecting students to citizen science and curated collections" module is an example of how a community of practice can work collaboratively to enhance training in foundational components of biology curricula, in this case for a plant collection and identification exercise, and to introduce data acumen in an authentic context.

## Acknowledgments

## Supplemental material
Supplemental data are available at *BIOSCI* online.

## References cited

[AAAS] American Association for the Advancement of Science. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action. AAAS.

Audubon Society. Why We Changed From "Citizen Science" to "Community Science." Audubon Center at Debs Park. www.audubon.org/about/edi.

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P. 2019. Research applications of primary biodiversity databases in the digital age. PLOS ONE 14: e0215794.

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: e1005755.

Beardsley TM, Gropp RE, Verdier JM. 2018. Addressing biological informatics workforce needs: A report from the AIBS Council. BioScience 68: 847–885.

Bik HM. 2017. Let's rise up to unite taxonomy and technology. PLOS Biology 15: e2002231.

Biró É, Babai D, Bódis J, Molnár Z. 2014. Lack of knowledge or loss of knowledge? Traditional ecological knowledge of population dynamics of threatened plant species in East–Central Europe. Journal for Nature Conservation 22: 318–325.

Biswas-Diener R, Jhangiani R. 2017. Open: The Philosophy and Practices that Are Revolutionizing Education and Science. Ubiquity Press.

Britz R, Hundsdörfer A, Fritz U. 2020. Funding, training, permits: The three big challenges of taxonomy. Megataxa 1: 049–052.

Carroll SR, et al. 2020. The CARE Principles for indigenous data governance. Data Science Journal 19: 43.

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

Cook JA, et al. 2020. Integrating biodiversity infrastructure into pathogen discovery and mitigation of emerging infectious diseases. BioScience 70: 531–534.

Crisci JV, Katinas L, Apodaca MJ, Hoch PC. 2020. The end of botany. Trends in Plant Science 25: 1173–1176.

Ellwood ER, et al. 2020. Biodiversity science and the twenty-first century workforce. BioScience 70: 119–121.

Farley SS, Dawson A, Goring SJ, Williams JW. 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. BioScience 68: 563–576.

Funk VA. 2018. Collections-based science in the 21st century. Journal of Systematics and Evolution 56: 175–193.

Garcia L, et al. 2020. Ten simple rules for making training materials FAIR. PLOS Computational Biology 16: e1007854.

Hackett RA, Belitz MW, Gilbert EE, Monfils AK. 2019. A data management workflow of biodiversity data from the field to data users. Applications in Plant Science 7: e11310.

Hampton SE, et al. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11: 156–162.

Hampton SE, et al. 2017. Skills and knowledge for data-intensive environmental research. BioScience 67: 546–557.

Hardisty A, Ma K, Nelson G, Fortes J. 2019. openDS: A new standard for digital specimens and other natural science digital object types. Biodiversity Information Science and Standards 3: e37033.

Heberling JM, Isaac BL. 2018. iNaturalist as a tool to expand the research value of museum specimens. Applications in Plant Science 6: e01193.

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.

Henkhaus N, et al. 2020. Plant science decadal vision 2020–2030: Reimagining the potential of plants for a healthy and sustainable future. Plant Direct 4: e00252. https://doi.org/10.1002/pld3.252

Hobern D, et al. 2019. Connecting data and expertise: A new alliance for biodiversity knowledge. Biodiversity Data Journal 7: e33679.

Indiana University. 2017. The Carnegie Classification of Institutions of Higher Education. About the Carnegie Classifications. Indiana University Center for Postsecondary Research. https://carnegieclassifications.iu.edu.

Jarić I, et al. 2020. iEcology: Harnessing large online resources to generate ecological insights. Trends in Ecology and Evolution 35: 630–639.

Lacey EA, et al. 2017. Climate change, collections and the classroom: Using big data to tackle big problems. Evolution: Education and Outreach 10: 1–13.

LaDeau SL, Han BA, Rosi-Marshall EJ, Weathers KC. 2017. The next decade of big data in ecosystem science. Ecosystems 20: 274–283.

Lendemer J, et al. 2020. The extended specimen network: A strategy to enhance us biodiversity collections, promote research and education. BioScience 70: 23–30.

Linton D, Ellwood E, Phillips M, Krimmel E, Douglas N, Monfils AK. 2020. Backward Design with Digital Data. Biodiversity Literacy in Undergraduate Education, QUBES Educational Resources. https://doi.org/10.25334/9SDG-5Z52

McLean BS, et al. 2016. Natural history collections-based research: Progress, promise, and best practices. Journal of Mammalogy 97: 287–297.

McNemar Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12: 153–157.

Middendorf G, Pohlad BR. 2014. Ecoliteracy for ecology and ecologists: Eroded underpinnings. Frontiers in Ecology and the Environment 12: 194–195.

Miller SE, et al. 2020. Building natural history collections for the twenty-first century and beyond. BioScience 70: 674–687.

Monfils AK, Powers KE, Marshall CJ, Martine CT, Smith JF, Prather LA. 2016. Natural history collections: Teaching about biodiversity across time, space, and digital platforms. Southeastern Naturalist 16: 47–57.

Monfils A, Krimmel E, Linton D, Marsico T, Morris A, Ruhfel B. 2019. Connecting Students to Citizen Science and Curated Collections. Biodiversity Literacy in Undergraduate Education, QUBES Educational Resources. https://doi.org/10.25334/Q4DN0M.

Monfils AK, et al. 2020. Regional collections are an essential component of biodiversity research infrastructure. BioScience 70: 1045–1047. https://doi.org/10.1093/biosci/biaa102.

[NASEM] National Academies of Sciences, Engineering, and Medicine. 2018. Data Science for Undergraduates: Opportunities and Options. National Academies Press. https://doi.org/10.17226/25104.

[NASEM] National Academies of Sciences, Engineering, and Medicine. 2020. Biological Collections: Ensuring Critical Research and Education for the 21st Century. National Academies Press. https://doi.org/10.17226/25592.

[NITRD] Networking and Information Technology Research and Development Program. 2016. The federal big data research and development strategic plan. NITRD. www.nitrd.gov/pubs/bigdatardstrategicplan.pdf.

Page LM, MacFadden BJ, Fortes JA, Soltis PS, Riccardi G. 2015. Digitization of biodiversity collections reveals biggest data on biodiversity. BioScience 65: 841–842.

Parsley KM. 2020. Plant awareness disparity: A case for renaming plant blindness. Plants People Planet 2: 598–601.

Powell C, Motley J, Qin H, Shaw J. 2019. A born-digital field-to-database solution for collections-based research using collNotes and collBook. Applications in Plant Sciences 7: e11284.

Powers KE, Prather LA, Cook JA, Woolley J, Bart HI, Monfils AK, Sierwald P. 2014. Revolutionizing the use of natural history collections in education. Science Education Review 13: 24–33.

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.

Raven PH, Miller SE. 2020. Here today, gone tomorrow. Science 370: 149.

Robeva RS, Jungck JR, Gross LJ. 2020. Changing the nature of quantitative biology education: Data science as a driver. Bulletin of Mathematical Biology 82: 127. https://doi.org/10.1007/s11538-020-00785-0.

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

Suarez AV, Tsutsui ND. 2004. The value of museum collections for research and society. BioScience 54: 66–74.

Turner H. 2015. Decolonizing ethnographic documentation: A critical history of the early museum catalogs at the Smithsonian's National Museum of Natural History. Cataloging and Classification Quarterly 53: 658–676.

Thompson CW, et al. 2021. Preserve a voucher specimen! The critical need for integrating natural history collections in infectious disease studies. mBio 12: e02698–02620.

Wandersee JH, Schussler EE. 1999. Preventing plant blindness. American Biology Teacher 61: 82–86.

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, vol. 50. CRC Press.

Wenger E. 2011. Communities of Practice: A Brief Introduction. University of Oregon. http://hdl.handle.net/1794/11736.

Wen J, Ickert-Bond SM, Appelhans MS, Dorr LJ, Funk VA. 2015. Collections-based systematics: Opportunities and outlook for 2050. Journal of Systematics and Evolution 53: 477–488.

Wiggins G, McTighe J. 2005. Understanding by Design, 2nd ed. Association for Supervision and Curriculum Development.

Wilcoxon F. 1992. Individual comparisons by ranking methods. Pages 196–202 in Kotz S, Johnson NL, eds. Breakthroughs in Statistics: Methodology and Distribution. Springer. https://doi.org/10.1007/978-1-4612-4380-9_16

Wilkinson MD, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018.

Wüest RO, et al. 2020. Macroecology in the age of big data: Where to go from here? Journal of Biogeography 47: 1–12.

*Anna K. Monfils (monfi1ak@cmich.edu) and Debra L. Linton are affiliated with Central Michigan University, in Mount Pleasant, Michigan, in the United States. Erica R. Krimmel is affiliated with Florida State University, in Tallahassee, Florida, in the United States. Travis D. Marsico is affiliated with Arkansas State University, in Jonesboro, Arkansas, in the United States. Ashley B. Morris is affiliated with Furman University, in Greenville, South Carolina, in the United States. Brad R. Ruhfel is affiliated with the University of Michigan, in Ann Arbor, Michigan, in the United States.*