The current era of global biotic change requires rapid identification of the onset of potentially adverse processes (e.g., arriving invasive species, phenological shifts) to facilitate effective countermeasures. These adverse processes may be initially marked by outliers (otherwise described as anomalies, oddities, peculiarities, abnormalities, and with other various terms): observations that fall outside of a previously understood norm. Early detection and rapid response systems focused on biotic change are quickly expanding beyond ecological monitoring (e.g., Anderson-Teixeira et al., 2014) and formal detection networks (e.g., Federal Interagency Committee for the Management of Noxious and Exotic Weeds, 2003) toward ever more creative solutions, such as using new technologies (e.g., NEON’s phenoCam network; Brown et al., 2016) or enlisting the help of indigenous populations (Lauer and Aswani, 2010), citizen scientists (Prince and Zuckerberg, 2014; Scyphers et al., 2015), and even social media users (Daume, 2016), to discover and track biotic change. Clearly, now is the time for creative solutions for recognizing change, but the mixed successes of unconventional detectors of
change (Lauer and Aswani, 2010) underscores the need to mobilize new groups of experienced professionals to step into this role. One such group well-suited to the task of observing, documenting, and reporting change are collectors and preparators of biodiversity specimens (henceforth referred to here as "collectors"), including botanists, entomologists, ornithologists, mammalogists, and experts on other taxonomic groups. However, we have found in a survey of collectors, a search of collector training materials, and an exploration of identified outlier terms in over 75 million specimen records, that the collector community is hampered by several factors. Here, we present our findings and seek to mobilize the community as more effective sentinels of change.

Collectors are a diverse group of researchers who collect organisms or parts of organisms for identification, study, documentation, and preservation in natural history collections, including herbaria. We began with the observation that collectors often have extensive personal experience, a network of scholarly resources, detective skills honed to recognize differences among organisms, and the motivation to resolve incongruities with existing resources when identifying their collections, especially since unique specimens may indicate new species. This last point is worth reinforcing: while many researchers might discard outliers (e.g., during statistical analyses), collectors are motivated to pay them special attention. Nevertheless, collectors’ observations of outliers appear to be regularly lost or, at most, noted on a label where the observation is left to the vagaries of future digitization and data mining. Indeed, most outlier detection occurs in an ad hoc, post hoc analysis of patterns among historically collected specimens, rather than using the collector’s observations at the time of collection (e.g., Peñuelas and Filella, 2002; Meyers et al., 2009; Ożgo and Schilthuizen, 2012; Bucher and Aramburu, 2014). Meanwhile, significant biotic change can occur over stunningly brief periods (Franks et al., 2007; Stuart et al., 2014), revealing a need for rapidity that traditional, retrospective methods of outlier detection lack. Leveraging the experience and in situ activity of collectors could become a transformative step for global change biology if collectors were encouraged and empowered to immediately report their observations to stakeholders—e.g., scientists, land managers, and policymakers who study, prevent, and manage biotic change—in an effective way that adequately presents the evidence for the conclusion.

Biotic change is a constant over long spatial and temporal scales (e.g., Gatti et al., 2015; Tennant et al., 2017), but the Anthropocene—a geological unit still without formal recognition, but likely to be recognized as beginning in the mid-20th century based on the May 2019 vote of the International Commission on Stratigraphy’s Anthropocene Working Group—has “rapid changes in the biosphere” among its several noted phenomena (AWG, 2019). Failure to identify the beginnings of adverse changes in the biosphere can have significant implications for biodiversity loss (Cardinale et al., 2012), costly exotic species invasions (U.S. Fish and Wildlife Service, 2012), and declines in ecosystem services such as pollination (e.g., Potts et al., 2010). We identified seven outlier types that could be apparent to collectors and for which they might have resources or personal experience with which to construct a norm: distributional (e.g., occurrence outside a previously understood geographic range), phenological (e.g., life history events outside a previously understood temporal range), ecological (e.g., new interactions in a food web), morphological (e.g., new color variants), anatomical (e.g., skeletal deformities), behavioral (e.g., new migration patterns), and genetic (e.g., previously undescribed hybrids; see question 3 of Appendix S1 for further examples). These outlier types are often interrelated due to cascading effects. For example, newly arrived species might lead to genetic outliers (e.g., hybrids) that are first revealed by morphological and behavioral changes, which may then produce altered ecological interactions.

Determining the exact causes of biological outliers is often beyond the purview of the collector and can be quite complicated to determine. For example, morphological, anatomical, and phenological outliers might be due to previously unexpressed phenotypic plasticity (Torres-Dowdall et al., 2012), physiological stress due to environmental change (Harper and Wolf, 2009), significant mutations such as polyploidy (Otto and Whitton, 2000), hybridization (López-Camal and Tovar-Sánchez, 2014), or the fingerprint of selective processes (Grant and Grant, 1995). If anything, this complexity argues for a smooth and timely hand-off of information between those who first notice the outlier and the stakeholders who use that information for future study, policy, and management. When properly empowered, biodiversity specimen collectors can play a critical role in quickly detecting and reporting biotic change, yet the collecting community remains largely untapped for this purpose. With our survey and analyses, we aim to illuminate impediments to effectively mobilizing this community and to catalyze future action.

COMMUNITY SURVEY AND ANALYSES

We designed an 18-question survey of biodiversity specimen collectors to determine (1) whether collectors detect outliers other than new taxa, (2) how and with what frequency collectors detect outliers, (3) how collectors document and report outliers, (4) what resources were employed during their training as collectors, and (5) what factors impede outlier documentation and reporting (Appendix S1). We distributed a web link and brief description of the survey via natural history collection listserves (ECN-L, HERBARIA, iDigBio, NHCOLL-L, SERNEC, TAXACOM, TDWG), iDigBio social media (Facebook, Twitter; iDigBio is the U. S. National Science Foundation’s National Resource for Advancing Digitization of Biodiversity Specimens), and with the help of the Natural Science Collections Alliance (NSCA). At least four professional societies within the NSCA also distributed the link through social media or email lists. The survey was open for 36 days, and we sent a reminder email via the same listserves 2 weeks before the close of the survey. All respondents gave their informed consent to participate online before completing the survey, and ethical approval for the survey and our study methods was obtained from the Institutional Review Board of the Florida State University (current record number 2019.26915) before distribution.

The survey received 222 responses with representation from collectors of 10 major groups of organisms and additional groups in the "Other" category. The greatest number of respondents collect plants (55 respondents; 25% of total), followed by those collecting insects (50; 23%), fungi (39; 18%), and algae (18; 8%). A wide range of collecting experience (0–31+ years) was represented among survey respondents, with a relatively even distribution of 36% falling into the 0–15 years range, 29% in the 16–30 years range, and 36% in the 31+ years. Not all participants answered every survey question, and each question received an average response rate of 82% (Appendix S1).

Collectors provided 170 unique words and phrases that they would use to describe outliers (Fig. 1); 70% of these were not
Collectors discover some outliers frequently and easily

Our survey results corroborate our main premise: collectors detect specimen outliers, and they do so using a rich suite of resources. Most frequently, collectors reference other specimens of the same taxon when identifying outliers (91% of respondents), but they also use personal experience (89%), monographs, and other highly vetted resources (82%), taxonomic experts (71%), online specimen data aggregators (53%), and other resources (Appendix S1).

About 80% of respondents (excluding responses of “Does Not Apply”) indicated that they discover distributional and morphological outliers at least “occasionally”, and almost half of respondents considered those two types of outlier “Easy” or “Very Easy” to detect. On the other hand, more than half of respondents reported that they rarely or never discover behavioral, genetic, and anatomical outliers, and more than half deemed them “Difficult” or “Very Difficult” to detect. Over half of respondents indicated discovering phenological outliers at least “occasionally”, and phenological outliers were considered similarly “Difficult” or “Very Difficult” (42%) and “Easy” or “Very Easy” (32%) to detect. These results may indicate important differences in detectability of certain outliers between taxonomic groups. For example, earlier-flowering trees may be more conspicuous than, for example, nocturnal animals that emerge from hibernation early.

But we identified relatively few outlier reports among the specimen records

More than half of respondents (59%) reported noting outliers on specimen labels, and half (50%) reported noting outliers in field notes/journals. Data from both of those sources might be expected to eventually appear in digital specimen records after some time lag. Few (16%) reported that they “generally do not make note of outliers”.

However, very few of the 75 million specimen records that we queried contained even the most frequently cited outlier terms. Survey respondents most frequently cited “abnormal” as used to document both morphological and anatomical outliers, “early” to document phenological outliers, “extension” to document distributional outliers, “unusual” to document ecological outliers, and “hybrid” to document genetic outliers. However, our queries of specimen records indicate that these terms may not be used in these contexts consistently. Of these terms, “hybrid” appeared most frequently in the iDigBio-aggregated data (in as many as 19,556 of 75 million records, about 0.03%), followed by “early” (16,218 after removing fields that were references to geological time periods; about 0.02%) and “extension” (13,177; about 0.02%). Other previously noted terms appeared much less frequently than 0.01% of the records. “Unusual” appeared in just 2862 field values, and “abnormal” repeated across the survey respondents. Some of these words alone could communicate outlier status (e.g., “atypical”, “strange”, and “aberrant”), whereas most would only do so in a longer phrase (e.g., “fruiting”, “small”, and “new”). We queried all available fields of the 75,569,035 specimen records (as of 30 January 2017) served by iDigBio, a major aggregator of biodiversity specimen records, for these “outlier terms” (some were phrases, such as “than normal”). We sought to determine their frequency and distribution of use in digitized specimen data with special attention given to the use of those words that alone could communicate outlier status and therefore support a stakeholder’s search of the data using existing search functionality of the iDigBio web portal and application programming interface (the latter typically shortened to API).

Collectors named 84 unique literature references used when learning to collect. Many of these, such as monographs and field guides, proved to be taxonomic resources rather than instructional materials and were thus considered out-of-scope. Only 22 of the provided references were considered in-scope, and seven of these were inaccessible to us. Starting with these 15 resources, we assembled additional collector training literature (e.g., institutional manuals and written protocols) from works cited, collection websites, and other sources (see Appendix S2 for methods and final bibliography). The search produced an additional 32 books, articles, protocols, manuals, and websites. We examined all materials for any reference to methods of detecting, documenting, or reporting biological outliers.

FIGURE 1. Word cloud of terms and phrases used by collectors to describe outliers, as reported in our community survey. Relative word sizes correspond to relative number of collectors who listed the term.
in just 623. Additional terms that could communicate outlier status without being contained in a longer phrase (including “odd,” “atypical,” “strange,” “vagrant,” “aberrant,” “weird,” “straggler,” and “unexpected”) also appeared less frequently than 0.01% of the records. Outlier, the term that we use here to describe our topic, appeared in 1085 field values, but most of these were references to the locality (e.g., “SW outlier of the Sierra de Manantlán”), rather than a quality of the specimen.

Furthermore, using the terms to search particular fields for specific outlier types is not straightforward. This challenge arises because some terms have multiple meanings (e.g., “early” to describe paleontology and geological time periods), some terms are used to describe phenomena that are common in some taxonomic groups (e.g., “hybrid” in plants) but not in others, some terms can flag different types of outliers (e.g., a specimen can have “abnormal” morphology, anatomy, behavior, etc.), all terms are found in many fields (Fig. 2), and most term-by-field combinations did not produce results of single outlier types. Terms that appeared in at least one database field were present in an average of 15 fields (median = 13); 32 terms did not appear in any records. Of the highly cited terms, “early” appeared in 38 fields (most frequently in occurrenceRemarks after earliestEpochOrLowestSeries), “hybrid” appeared in 32 fields (most frequently in scientificName), “unusual” in 25 (most frequently in occurrenceRemarks), “abnormal” in 17 fields (most frequently in occurrenceRemarks), and “extension” in 17 (most frequently in dynamicProperties). Other seemingly promising terms recognized above appear most frequently in occurrenceRemarks except “strange” and outlier (which occur most frequently in locality; e.g., “Strange Rd.” and “SW outlier”), “vagrant” (most frequently in habitat), “straggler” (most frequently in taxonRemarks), and “unexpected” (most frequently in fieldNotes).

Of note, many of these fields are not high priorities in digitization activities designed to produce skeletal data records at high rates, which might partly explain the lower frequency of outlier terms in records than expected. In other words, outlier information might fall out-of-scope in first-pass digitization of specimens, and their absence in data records is due to the decisions made at the digitization step by project managers rather than at the label-creation step by collectors. The degree to which this affects conclusions drawn from the iDigBio-aggregated records is unclear, but even if half of the records were incomplete in this way, the frequency of outlier terms would not exceed 6 in 10,000 (for “hybrid”, the most frequently observed outlier term), a frequency that remains strikingly low. A further potential concern—that

![FIGURE 2. Heat map representing the frequency of terms provided by survey respondents to describe outliers (y-axis) in data fields (x-axis) of 75,569,035 specimen records served by iDigBio on 30 January 2017. Only the terms provided by at least two survey respondents and only the 50 data fields containing the highest frequency of outlier term occurrences are shown for simplicity. Dark red values indicate high frequency of records containing a given term (maximum: 218,250 instances); grey values indicate low frequency of records (minimum: 0 instances). Values were log-transformed for ease of viewing.](image-url)
project managers are selecting to digitize type specimens, which might rarely use outlier terms, instead of a representative sample of their general collections—is more easily addressed. Less than 1% of iDigBio records are noted to be type specimens (as of 1 April 2019), suggesting that they are influencing the conclusions in, at most, small ways.

Impediments to reporting are diverse but include lack of protocols and training

What might be impeding outlier reporting, as is suggested by the rarity of records with outlier terms? About a quarter of respondents (26%) reported not generally being impeded in noting or reporting outliers, suggesting that three-quarters of respondents to the question recognize impediments. The greatest number of respondents reported that lack of time served as an impediment (47%), followed by low confidence in the outlier status (43%), and lack of standard community protocols for reporting (30%).

Half of survey participants indicated that they were not taught or were self-taught to document outliers. About 17% indicated they were taught on the job or by the example of mentors or other professionals. Only 3% of participants indicated that they learned to document outliers by reading literature, and just 2% indicated that they learned through courses. Accordingly, only one of the 47 training references we searched (Resources Inventory Committee, 1999) included recommendations for documenting outliers or anomalies. None described the important role of collectors in documenting change or provided best practice recommendations for the activity.

The community seems receptive to mobilization as sentinels of change

Our study suggests that collectors should be an easy group to mobilize for outlier detection and documentation. A large percentage (80%) of collectors view themselves as at least somewhat being “on the front lines of observing and documenting change in Earth’s biota”, and the majority of respondents (59%) have seen their attitudes on the topic change at least somewhat during their career. Given the latter, we were somewhat surprised that we did not find evidence that outlier terms were becoming more frequently used (A. R. Mast and K. D. Pearson, unpublished results). For example, the percentages of records that used “unusual” and “odd” peaked in the 1990s and have since declined.

Many collectors report outliers other than new taxa to close colleagues (74%), experts on the respective taxon (65%), and in journal articles or reports (52%), though fewer do so to other potential stakeholders (e.g., land managers or park staff, 27%; government agencies, 21%; and enthusiast groups, 14%) or to online community resources that those stakeholders might use (e.g., iNaturalist or EDDMapS, 9%).

MOVING FORWARD

We seek to make outlier detection, documentation, and reporting core activities of the collection community, and we view community buy-in and collaborative progress as critical to moving beyond what might be described as abundant but unfocused goodwill toward the activities. We recognize three steps in the collector community’s outlier-related data flow that merit further attention given our findings (Fig. 3): (1) immediate reporting to a broader set of stakeholders, (2) reporting on the specimen label, and (3) digitization of specimen label data. Our study identified a near-complete absence of training resources related to these steps. We suggest that a first step will be to develop best practices for each in such settings as workshops and working groups, at which times requirements of missing resources such as data standards and cyberinfrastructure can additionally be addressed.

Best practice recommendations are outside the scope of our current paper because we view them as best developed in more broadly representative groups, but we will make a few observations that could be useful when framing those community-driven activities. First, we recognize that if collection data were “born digital,” as is likely in the future, Fig. 3 would look different. That is, were collection data regularly entered directly into a specimen data management system by the collector (e.g., via an application [app] on a mobile device) with subsequent creation of the specimen label from that entry, then best practices for outlier description could be supported in the data entry forms and publication of the outlier information automated with the specimen data management system as the provider. The three numbered steps in the figure could then

FIGURE 3. Simplified model of the process of outlier detection, documentation, and reporting by collectors. The steps we identified as needing further attention are highlighted by red numbers: (1) immediate reporting to a broader set of stakeholders, (2) documenting on the specimen label, and (3) digitizing specimen label data. Our survey suggests that few collectors report outliers directly to stakeholders (e.g., land managers), which is reflected by the dashed arrow. See text for additional explanation.
occur simultaneously or in close succession. The majority of data produced by the community reporting observations (that is, occurrence data not anchored to a specimen, such as at eBird.org and iNaturalist.org) is "born digital" and represents a potential model and partner for methods and infrastructure development. For example, eBird's mobile app provides users with information about which birds are rarely seen in their current location and thus would represent a distributional outlier before submitting the observation. And both iNaturalist and eBird have active forums in which users can immediately discuss the accuracy and significance of outlier observations.

Although data collection methods might change, it is useful to recognize the three steps in Fig. 3 since they reflect the current procedure of most collectors, and each highlights a different challenge for the community to address. Step (1) might involve the construction or tailoring of a subscription service for outlier observations, which has parallels with trailblazing projects like FilteredPush (for digital annotations of specimen data; Wang et al., 2009) and FreshData (for occurrence data more generally; Hammock and Polen, 2016). Ideally, this step would result in the data efficiently arriving in the "go-to" sites that stakeholders might visit to find relevant types of outlier data, such as the Early Detection and Distribution Mapping System (Wallace and Barger, 2014; eddmaps.org) for invasive species and National Phenology Network (Schwartz et al., 2012; usanpn.org) for phenology. Step (2) involves greater semantic sophistication in the description of outlier observations, which should take into account data discovery mechanisms (e.g., at iDigBio and the Global Biodiversity Information Facility [GBIF]), data models at the go-to sites for stakeholders like those mentioned above, and work on relevant ontologies, such as the Plant Phenology Ontology (Stucky et al., 2017; Willis et al., 2017) and the Open Annotation Ontology (e.g., by Morris et al., 2013). Step (3) involves careful consideration of the downstream discovery and reuse of the collector's observation, enabling data aggregators (e.g., iDigBio and GBIF) to optimize discovery and delivery of this type of data, and needs to engage the deep expertise on the topic represented by the Biodiversity Information Standards (TDWG) group. Future work on these steps will also benefit from the foundation that our community survey results and subsequent analyses provide.

In the meantime, we encourage collectors to report outliers in timely and effective ways to broad audiences of potential stakeholders appropriate to the outlier type, taxon, and location and then tell the collector community how they did it. We encourage that special care be taken to adequately document the evidence for a specimen's outlier status in these reports to stakeholders in ways that clarify the outlier assessment's fitness-for-use and maintains trust in the collector community. Twitter is an increasingly effective, low-barrier mechanism for community conversations about protocols because the communications are open to everyone, and we suggest tweeting the information with the hashtag #ODDbyCollectors (shorthand for Outlier Detection and Documentation by Collectors). We started the conversation by tweeting an observation on this topic, typically involving a concrete example of an outlier specimen, most weeks in 2018, and these can be found by searching Twitter for that hashtag. We invite you to join this conversation and encourage those with a deeper interest to join the iDigBio Outlier Detection and Documentation by Collectors Working Group that we co-chair.

Collectors are ideally suited to provide early warnings of biotic change in the Anthropocene for the benefit of science and society. Let's work together to ensure that our community is successfully mobilized for this purpose.

ACKNOWLEDGEMENTS

The authors thank Kevin Love, Alex Thompson, and Matthew Collins for assistance in producing data for Fig. 2; Shari Ellis for management of the survey logistics; and Libby Ellwood, Gil Nelson, Deb Paul, and Greg Riccardi for support and insight. We thank Andrew Kratter, Gil Nelson, Max Nickerson, Larry Page, and David Reed for introducing us to collection and digitization practices in their respective taxonomic domains. We thank those who distributed the survey, including member societies of the Natural History Collections Alliance. We thank all who participated in the survey. And we thank two anonymous reviewers and an anonymous Associate Editor for suggestions.

This research was supported through iDigBio, which is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (award number 1547229). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

AUTHOR CONTRIBUTIONS

K.D.P. and A.R.M. conceptualized, designed, and conducted the survey and analyzed survey results. K.D.P. conducted the literature survey and A.R.M. summarized iDigBio results. K.D.P. and A.R.M. wrote and edited the final manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Results of survey of biodiversity specimen collectors and preparators.

APPENDIX S2. Methods and bibliography of survey of training materials used to learn collecting methods and protocols.

LITERATURE CITED

Anderson-Teixeira, K. J., S. J. Davies, A. C. Bennett, E. B. Gonzalez-Akre, H. C. Muller-Landau, S. J. Wright, K. A. Salim, et al. 2014. CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. Global Change Biology 21: 528–549.

Brown, T. B., K. R. Hultine, H. Steltzer, E. G. Denny, M. W. Denslow, J. Granados, S. Henderson, et al. 2016. Using phenocams to monitor our changing Earth: toward a global phenocam network. Frontiers in Ecology and the Environment 14: 84–93.

Bucher, E. H., and R. M. Aramburu. 2014. Land-use changes and monk parakeet expansion in the pampas grasslands of Argentina. Journal of Biogeography 41: 1160–1170.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, et al. 2012. Biodiversity loss and its impact on humanity. Nature 486: 59–67.

Daume, S. 2016. Mining Twitter to monitor invasive alien species—an analytical framework and sample information topologies. Ecological Informatics 31: 70–82.

Federal Interagency Committee for the Management of Noxious and Exotic Weeds. 2003. A national early detection and rapid response system for...
