Singer Randal (Orcid ID: 0000-0001-9927-1748)

REGULAR PAPER

Awareness and Use of Biodiversity Collections by Fish Biologists

Randal Singer[1] | Shari Ellis[2] | Lawrence M. Page[2]

[1] University of Michigan Museum of Zoology, University of Michigan, Ann Arbor, Michigan, USA

[2] Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA

Correspondence

Randal Singer, University of Michigan Museum of Zoology, University of Michigan, 3600 Varsity Drive, Ann Arbor, MI 48108 , USA

Email: rsinger@umich.edu

Funding information

This project was supported by iDigBio with funding from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (EF1115210, DBI 1547229).

Abstract

A survey of 280 fish biologists from a diverse pool of disciplines was conducted in order to assess the use made of biodiversity collections and how collections can better collect, curate and share the data they have. From the responses, data for how fish biologists use collections, what data they find the most useful, what factors influence the decisions to use collections, how they access the data and explore why some fish biologists make the decision to not use biodiversity collections is collated and reported. The results of which could be used to formulate sustainability plans for collections administrators and staff who curate fish biodiversity collections, while also highlighting the diversity of data and uses to researchers.

KEYWORDS

best practice, biodiversity collections, natural history, survey, sustainability

1 | INTRODUCTION

Current understanding of how fish biologists use biodiversity collection data for research is largely inferred from reviews of published literature, analyses of institutional collection records, or a combination of the two (Abrahamson, 2015; Ball-Damerow *et al*., 2019; Bradley *et al*., 2014; Carrine *et al*., 2018; McClean *et al*., 2015). While these approaches provide insight into the use of collections, they fail to capture all uses of collections as researchers often neglect to adequately cite use of collections data in published reports (Ball-Damerow *et al*., 2019; Mooney and Newton, 2012) and some kinds of publications (*e.g.*, government reports) are not readily discovered using existing search tools.

Reviews of publications and analyses of institutional records also provide little insight into researchers' preferences regarding the kind of collections data researchers would like to access. Improving information about who uses natural history collections and why is critical to the sustainability of institutional collections (Page *et al.*, 2015). Reaching out to the scientists who could potentially use biodiversity collections but currently do not will lead to a better understanding of why some choose to use collections and others do not, the kinds of data valued and the questions collections are currently being used to answer.

We examined how fish collections are used by researchers as an example of the broad use of collections data. Fish collections were selected for study because of our personal interest and the availability of data from our previous analysis of fish collections (Singer, *et al*., 2018). Fish collections are used by a diverse array of scientific disciplines, including studies of public health (Barber *et al*., 1972; Miller *et al*., 1972), anatomy and physiology (Sparks *et al*., 2014), ecology (Pyke and Ehrlich, 2010) and conservation (Shaffer *et al*., 1998; Ponder *et al*., 2001). In this study, we surveyed fish biologists to explore the backgrounds of those who do and do not use fish collections in research, the types of data they find most valuable and how they access natural history collections data. These data were collected with the goal of informing the fish collections community and, when appropriate, the broader collections community about the needs and practices of researchers.

2 | MATERIALS AND METHODS

2.1 | Survey development

We constructed an online survey to collect information on use of fish collections, data preferences, research interests and background characteristics (*e.g.*, research specialty, training, collections experience and age) that could potentially influence decisions to use fish collections in research (Supporting Information). Questions about disciplines within fish biology, research interests and types of data were informed by the first and third authors' combined 50+ years' experience using and managing fish collections. Decisions about survey length, structure and question formats were based on recommendations provided by Dillman *et al*. (2014).

The survey was initially validated by having five fish collection managers and four other fish biologists complete the online survey to provide feedback about question clarity, appropriateness and format. Based on feedback, revisions were made resulting in a final version of the survey. Changes included adding an option for respondents to rate the usefulness of types of data beyond just their top five and some restructuring of how some questions were written in order to give more clarity. The final survey was approved by the University of Florida's Institutional Review Board prior to distribution. The final version of the survey can be found in the online Supporting Information.

2.2 | Survey distribution

To ensure the online survey reached a wide range of fish biologists, we contacted 13 organisations known to have members with interests related to fish biology to inquire whether an anonymous link to our survey could be sent *via* the organisations' Listservs (*e.g.*, Natural History Collections Listserver or NHCOLL-L), other email services, or Facebook pages (*e.g.*, Ichthyology page). Eleven professional societies agreed while two (Ecological Society of America and Association for the Sciences of Limnology and Oceanography) declined because of privacy policies. In addition, the first author emailed invitations with the link to his personal network and posted a link to the survey on his personal Facebook page and on Twitter in addition to the Twitter hashtag #fishsci (Table 1). The survey was distributed on 1 March 2018 *via* the Qualtrics (2017) software licensed to the University of Florida. The survey was open for 3 months and closed 1 June 2018. We estimate the survey was viewed up to 29,400 times, based on the numbers obtained from organisations' business meeting minutes, social media analytics and personal communications with officers in the organizations contacted.

2.3 | Data Cleaning and Analysis

Three hundred and eighty-three respondents agreed to participate in the survey, but 103 of those individuals only answered the first two questions about their areas of work and published research. The latter were eliminated, leaving a convenient sample of 280 responses.

To facilitate analyses, respondents were sorted into discipline groups based on their patterns of responses when asked to select up to three areas in which they worked. Respondents were manually sorted into a discipline group based on the selection of these three disciplines. The groups were then named, based on the predominant discipline(s) identified by the respondents. These patterns were then compared to a computational cluster analysis of the survey responses for discipline of interest using JMP 14.3 (www.jmp.com). These data were used to confirm the groupings identified through manual analysis of the data. In all cases the results from the computational cluster analysis corroborated the manual clusters.

3 | RESULTS

3.1 | Respondents' demographics and field of work

Of the 280 respondents, 248 provided information on their age, educational background and employment. The modal age range was 25–34 years, with 42% of respondents falling within that age group. The other age groups were: 18–24 years (4%), 35–44 years (23%), 45–54 years (14%), 55–64 years (12%), 65–74 years (4%) and 75 years and older (1%). Most respondents had a doctorate (48%) or a master's degree (36%) with 14% having a bachelor's degree as their sole degree. The remaining 2% held a high school degree–general education diploma, associate of arts, or other degree. Most respondents (62%) were affiliated with an institution of higher education, including two-year community colleges (0.5%) or four-year universities (61.5%). The remaining respondents worked in state or federal government (23%), non-profit or business sectors (6%), or were retired (2%). The remaining respondents (5%) selected 'other' and specified no work organisation.

Respondents were asked to identify up to three areas in fish biology in which they work from a dropdown menu of 14 choices: ecology, behavioral ecology, conservation biology, fisheries management, taxonomy, phylogenetics, biogeography, morphology–functional anatomy, palaeontology, aquaculture, kinematics, epidemiology–disease ecology, microbiology and other. Of the 'other' responses, three were of sufficient frequency (at least 21 responses) to be included in the analyses: evolution, physiology and parasitology, bringing the total number of disciplines to 17. Although respondents were asked to limit their selections to three, the survey software did not prevent them from selecting more than three. The number of areas selected ranged from zero to nine, with an average of two. Ecology (n = 147) was by far the most frequently selected area followed by conservation biology (n = 104; Figure 1).

From the respondents who identified disciplines of interest, eight discipline groups were identified along with the relative rate of response for each (Table 2). Eleven individuals were excluded from this analysis because they selected too many disciplines to confidently assign them to a group. While the groups varied in terms of the number of disciplines selected, there were clear patterns in terms of the clusters of disciplines selected by most individuals assigned to each group. For example, those classified in fisheries management–conservation, all indicated fisheries management as a discipline of interest as well as ecology, conservation, or both. A third of the individuals in this group selected a fourth area of interest, but none of these areas reached a criterion of at least 50% of respondents in that group. In contrast, of the 73 individuals assigned to the multidisciplinary ecology group, 54 selected at least one other area of interest; no respondent in this group selected either conservation or fisheries management, and none of the other disciplines reached a criterion of 50% of respondents.

The majority of respondents were active researchers with 80% reporting that they had published at least one article related to fish biology in the previous 4 years. Most respondents also had experience using specimens or collections data in their research with 62% reporting they currently use specimens or collections data, 20% reporting they had in the past, with only 15% reporting they did not use specimens or collections data and 3% not indicating either way.

3.2 | Collections use by professional title

Respondents currently using collections were asked to select a job title that best described their current position from a list of 11 options plus ‘other’. The only repeated positions in other were post-doctoral and private researcher, so these were added for a total of 14 titles (Figure 2).

Professional staff and university faculty were those most likely to be currently using collections (33% and 28%, respectively), followed by graduate students (27%) and post-doctoral scholars (8%).

3.3 | Useful types of data

Respondents were given a list of 26 types of data commonly found in fish biodiversity collections, plus the option to select other, with a field to enter the data type. They were then asked to identify the five they found most useful. By far, the types of data rated most useful were those most closely related to a catalogued specimen: accurate taxonomic identification and locality, followed by meristic or morphometric data, tissue samples and fluid-preserved specimens (Figure 3). Types of data marked lowest for the top five 'most useful' were audio recordings (very few fishes make sounds), isolated reproductive organs and isolated pharyngeal teeth (the last two can usually be observed *in situ*), fossils and associated parasites (few palaeontologists and parasitologists participated in this study). Histology slides were also marked low as a valuable data type. There were not a significant number of data types listed in other to add to the results. During early distribution of the survey, respondents expressed concern about being limited to rating the importance of only five types of data. The survey was adjusted so that respondents could rate the importance of the remaining types of data beyond their top five. The types of data rated as useful but not ranked in a respondent's top five most useful are shown in

Figure 4. Among the types of data of secondary importance, field notes (38%), environmental data (32%), early life stages of fishes (30%), live, colour photos (30%) and sex of specimens (30%) were most frequently selected.

We further explored whether preferences for types of data varied across discipline groups. As observed when looking at all respondents, the availability of a physical specimen was highly valued regardless of discipline group, but the specific types of data were more varied. Accurate taxonomic identification and locality information remained the top-rated data among most groups (Figure 5). Systematics (no morphology) did not include meristic or morphometric data or tissue samples in their top five, although these types of data were rated as very useful by the respondents overall. Systematics (no morphology) were also the only discipline group to include live, colour photos among the types of most useful data. The importance assigned to environmental data also varied by group. It was rated as most useful by systematics (no morphology), pure fisheries management and pure morphology, but not by the other five groups. After selecting their top five most useful types of data, respondents were asked to select up to five additional types of data that did not include the types identified in their top five (Figure 6). The majority of discipline groups valued environmental data, early life stages of fishes, field notes, live, colour photos and sex of specimen.

It is important to note that when looking at the data for all respondents and the breakdown by discipline groups, that the other lower ranking types of data are useful. Even though these types of data were ranked lower than others, they still address the needs of researchers but cater to a smaller community of researchers. Some researchers who could be interested in these types of data may not be aware of such resources or were not trained to utilise collections data.

3.4 | Past Exposure to Collections

Respondents were presented a list of 13 ways in which they may have interacted with collections in the past and results compared between current users and non-users (Figure 7). Most respondents who have used collections reported visiting collections as part of graduate research or working in a collection as a graduate student; nearly half reported having visited as part of undergraduate research or during an undergraduate course field trip. Those who have not used collections in research were far less likely to have any of these exposures to collections.

From the responses related to what could be described as educational experiences with collections, there was a relatively clear picture of how respondents had interacted with collections in the past. A significant proportion of respondents had interactions with collections during their undergraduate education through either a class visit, for undergraduate research, or for a collections-based job. Taking into account the small percentage that had visited or worked in a collection during their primary and secondary, most respondents had been exposed to collections prior to starting their post-graduate (completion of masters or PhD) career. When looking at post-undergraduate (completion of a 2 or 4 year degree) education interactions, there was a large proportion of current collection users who had experience visiting or working in collections during graduate school. Graduate-level course visits, graduate research visits, collections-based jobs as a student or obtaining a position in a biodiversity collection accounted for 54% of the respondents' interactions with collections after completing graduate school (masters or PhD). Twenty-six per cent of respondents who were not currently using collections indicated that they had either never visited or never heard of biodiversity collections. In addition, very few of these respondents had hands-on experience working in collections.

3.5 | Reasons for use or non-use of collections data

Respondents who indicated they had never used collections in their research, or once had but were no longer, were presented a list of 10 potential reasons for not using collections as well as an opportunity to provide additional ones; respondents could select more than one reason. Respondents who had once used collections but were not currently using them identified, on average, one type of explanation for their lack of use, while those who never had used collections cited an average of 1.5 reasons. The responses were summarised into four categories as reasons for not using collections: lack of relevancy, access, data quality and job status or responsibilities (Figure 8). For those that had never used collections, lack of data relevancy was selected the most (52%), followed by lack of access (44%). For those that have discontinued their use of collections lack of relevancy was also the most identified reason (76%) followed by a change in job or research interest (38%).

3.6 | Methods by which users access collections

Most of the researchers currently using collections who answered the question (n = 174) reported using multiple means to access collections (mea$n$ = 3.8; range = 1–7). Downloading data from an institutional collection database was the most common (74%), followed by using data from a collection at their institution (68%), visiting a collection outside their institution (67%), borrowing specimens from a collection outside their institution (65%) and requesting specimen data *via* request to a curator or collection manager (63%). Fewer than half (45%) indicated they used data aggregators. Those who reported using collections in the past (n = 47) recalled using an average of 2.2 methods (range = 1–6). Downloading data from an institutional collection database was the most common (47%) followed by requesting data *via* a curator or collection manager (43%), with visiting a collection outside their institution (40%) being the next most frequently identified method. Borrowing specimens from a collection outside their institution (34%) and using a data aggregator (28%) were the least frequent methods for past access.

3.7 Use of data aggregators

As data aggregators are considered to be an effective way of sharing collections data (Page, *et al*. 2015) but were the least commonly used resource by fish biologists, a breakdown of aggregator use was made to examine this discrepancy. Respondents were shown the question regardless of collection use, because some researchers use aggregated data and still might not identify as a collection user. From 213 respondents, 65% indicated that they had used a data aggregator, while 35% reported that they had not. To explore which aggregators fish biologists preferred, respondents were given the following options: Global Biodiversity Information Facility (GBIF; www.gbif.org), iDigBio (www.idigbio.org), FishNet2 (www.fishnet2.net), VertNet (www.vertnet.org), Atlas of Living Australia (ALA; www.ala.org.au) and 'other'. There were not enough responses in other to warrant the addition of fields and most responses were one of the listed aggregators or would not be classified as a data aggregator (*e.g*., an institutional collection database). FishNet2 was by far the most selected (75%), with GBIF being the second at 50%. VertNet and iDigBio had equal usage with *c*. 30% and ALA was the least with *c*. 15%.

4 | DISCUSSION

Eighty per cent of the respondents to our survey reported that they had published at least one article in the previous 4 years and 82% reported having used data from fish collections in their research. These results provide insights into who uses data from natural-history collections in their research, the types of data they find most valuable and how the data are accessed. This information can inform the institutional collections community about the needs of researchers and identify improvements in data management that are likely to be of greatest value.

Ecology and conservation biology were the most frequently identified areas of research interest, followed by taxonomy and phylogenetics. Among the 26 types of data identified as commonly available in fish collections, those chosen as most valuable were accurate taxonomic identification, locality, morphological data, tissue samples and fluid-preserved specimens. The selection of accurate taxonomic identification is not surprising, given that respondents reported that downloading data from an institutional collection database was the most common method of accessing data and searches of data are commonly made on scientific names (Guala, 2016; König, *et al*. 2019). Also, publications and other reports on results of research almost always provide the scientific names of species studied and authors assume names provided in databases are correct. This indicates that a primary activity of collections staff should be keeping taxonomy current.

Selection of locality data as highly valuable also was to be expected given that ecologists and conservationists rely on accurate locality data to map distributions and assess demographic variation across landscapes (Carpenter, *et al*., 1993, Fernandez, *et al*., 2009) and taxonomists look at genetic and phenotypic variation throughout the geographic distributions of taxa when assigning names to populations (Winston, 1999). Almost all specimens arrive in collections with locality data, although collections staff sometimes need to add missing geographic coordinates. Legacy collections often lack coordinates and georeferencing has become a major activity of collections staff who use platforms such as GEOLocate to georeference large numbers of localities (Ellwood, *et al*., 2016; Seltmann, *et al*., 2018). Having all localities in a collection geo-referenced should be a goal of all institutions housing natural history collections.

The preference for morphological data and preserved specimens from which additional data, including those used in studies of phenology, can be acquired reflects the heavy reliance of phenotypic data in ecology and systematics. The value of tissue samples is likely to increase even beyond what it is currently given that phylogenetic analyses now rely overwhelmingly on genetic data generated from tissue samples preserved when specimens are collected (LaSalle, *et al*., 2016; Watanabe, 2019). Having tissues linked with specimens in collections enables researchers to depend on institutional collections for the identification of the species and geographic origin of the samples they use as well as allowing subsequent verification of the taxonomy if the genetic data suggest misidentifications.

For most respondents who use collections data in research, their awareness of the types of data and potential use in research were a result of exposure to collections as part of their college education. Many curators and other collections staff are involved in college teaching and outreach activities that provide excellent opportunities for increasing the awareness of collections as a source of data (Monfils, *et al*., 2017). However, given the effect these exposures have, an increase in these activities is likely to have significant benefits (Lacey, *et al*., 2017). Involvement in museum studies programs (WWL, 2019) that specifically concentrate on collections-based resources should also be a goal of collections staff. This could be done by directly participating or indirectly by providing information on specimens, uses of collections data, posters and other materials, or through the production of tutorials on collection management and data use.

Most respondents reported searching and downloading data directly from institutional databases rather from data aggregators. Fewer than half of the respondents indicated they accessed data through aggregators even though aggregation across institutions is a national focus in the USA (Pennisi, 2019) and clearly provides a more efficient method of data retrieval than individually visiting institutions or accessing individual institutional portals. The preference for institutional databases may be because they sometimes provide information not available from aggregators (*e.g.*, updated taxonomy, direct links to tissue samples, more images) or lead to efficiencies not characteristic of aggregators (Franz and Sterner, 2018; Poisot, *et al*., 2019). Most active researchers probably are aware of institutions that hold the most valuable information for their research (*e.g.*, because of a history of focused collecting in a particular geographic area by a curator) and consider it more efficient to go directly to these sources. This view is supported by the finding that more respondents rely on data obtained by directly contacting a curator or collection manager (43%), visiting a collection (40%), or borrowing and examining specimens (34%), than searching data aggregators. Among the aggregators most relevant to fish biologists, FishNet2 is most frequently used by respondents, presumably because of the (perceived or real) efficiency of searching among data only on fishes.

Results from our survey identified several areas where changes in collection management could increase the use of collections data in research. Given that downloading from a database was the preferred method of data retrieval for users of collections, it is important for collections to continue emphasising digitisation and data publishing. Having all localities in US collections geo-referenced is the goal of the Advancing Digitization of Biodiversity Collections (ADBC, National Science Foundation; www.nsf.gov) programme. As discussed above, it also is important that collections keep taxonomic and locality data as up-to-date and accurate as possible and the online access should reflect this current relevance.

It was evident that fish biologists require a wide variety of data types for their research. By increasing the diversity of data that collections curate and highlighting the precision and care that go into such curation, collections staff can reach more potential users and increase relevance to a broader pool of fish biologists. Some of this can be addressed with a more interdisciplinary approach during fieldwork with collectors gathering as much data pertaining to the specimen and its environment as time and resources permit. This information could include field notes, environmental metrics, early life stages of fishes (more attention to eggs and larvae) and colour photographs of live fishes. Involvement in museum studies programmes and university classes that specifically concentrate on collection-based resources should also be a goal of collections staff. Initiatives such as the RCN-UBE Biodiversity Literacy in Undergraduate Education (BLUE; www.biodiversityliteracy.com/) and Advancing Integration of Museums into Undergraduate Programs (AIM-UP; www.aimup.unm.edu/; Lacey, *et al.*, 2017) are steps in the right direction, but more community participation is needed to increase exposure to collection resources. Finally, it would be useful for websites of institutional collections and data aggregators to offer online tutorials to help potential users orient themselves with how to access and use the data for their research (Renaut, *et al*. 2018).

Despite the substantial sample of fish biologists that participated in the survey, there are limitations in the interpretation of the results. First, many data were collected with respect to why people use collections and what data types they find useful, but data on why people do not use collections could be more robust. In the future, focusing efforts on those fish biologists not using collections would be helpful in making recommendations that are more comprehensive and indicative of the collections community. In addition, 42% of the respondents were aged 25–34 while the rest of the respondents fell into smaller groupings. While this could mean that there are more individuals who identify as fish biologists in this age range, it might mean that this age group was simply more likely to participate in a digital survey. In the future, it might be worth addressing methods (other than digital surveys) to engage more researchers outside the 25–34 age range. Regardless, since most of the participants were in the age range of graduate students and younger professionals, these results might be particularly informative in relation to the next generation of fish biologists. There are clearly changes that could be made by collections staff to have large improvements on exposure, use and perceived value across the many scientific disciplines who study all aspects of fish biology. These changes can also be mirrored across the broader collections community to have an overall positive effect on collections sustainability and use.

ACKNOWLEDGEMENTS

We thank the many fish biologists who took the time to complete the survey. We especially thank the fish collection curators and collection managers who provided feedback and direction for what topics to address in the survey design, especially Z. Martin, B. Sidlauskas, R. Robins and M. Sabaj for detailed comments on survey design. We thank the many professional societies, clubs and social media groups who allowed for the dissemination of the survey link to the thousands of fish biologists across the world. Finally, we thank the University of Florida Statistics-Consulting Center for providing assistance with the analysis of discipline groups.

REFERENCES

Abrahamson, B. L. (2015) Tracking changes in natural history collections utilization: A case study at the Museum of Southwestern Biology at the University of New Mexico. *Collection Forum,* 29, 1-21.

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., LaFrance, R., Ariño, A. & Guralnick, R. (2019). Research applications of primary biodiversity databases in the digital age. *BioRxiv*, 605071.

Barber R. T., Vijayakumar A., & Cross F. A. (1972) Mercury concentrations in recent and ninety-year-old benthopelagic fish. *Science*, 178, 636-639.

Bradley, R. D., Bradley, L. C., Garner, H. J., & Baker, R. J. (2014) Assessing the value of natural history collections and addressing issues regarding long-term growth and care. *Biocience*, 64, 1150-1158.

Carine, M. A., Cesar, E. A., Ellis, L., Hunnex, J., Paul, A. M., Prakash, R., Rumsey, F. J., Wajer, J., Wilbraham, J., & Yesilyurt, J. C. (2018) Examining the spectra of herbarium uses and users. *Botany Letters*, 165, 3-4, 328-336.

Carpenter, G., Gillison, A., & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation, 2, 667–680.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail and mixed-mode surveys: the tailored design method*. John Wiley & Sons.

Ellwood, E. R., Bart, H. L., Jr., Doosey, M. H., Jue, D. K., Mann, J., Nelson, G., N. Rios, N., & Mast A. R. (2016) Mapping life – quality assessment of novice vs. expert georeferencers. Citizen Science: Theory and Practice 1, 4, 1–12. doi.org/10.5334/cstp.30

Fernandez, M.A., Blum, S. D., Reichle, S., Guo, Q., Holzman, B., & Hamilton, H. (2009) Locality uncertainty and the differential performance of four common niche-based modeling techniques. Biodiversity Informatics, 6, 36–52.

Franz, N. M., & Sterner, B. W. (2018). To increase trust, change the social design behind aggregated biodiversity data. Database, 2018, 1–12. doi.org/10.1093/database/bax100

Guala, G. F. (2016) The importance of scientific name synonyms in literature searches. PLoS One 11 (9): e0162648. doi.org/10.1371/journal.pone.0162648

JMP, Version 14.3. SAS Institute Inc., Cary, NC, 1989-2019.

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., & Kreft, H. (2019) Biodiversity data integration—the significance of data resolution and domain. PLoS Biol, 17 (3): e3000183. doi.org/10.1371/journal.pbio.3000183

La Salle, J., Williams, K.J., & Moritz, C. (2016) Biodiversity analysis in the digital era. Philosophical Transactions of the Royal Society B, 371 (1702): 20150337. doi.org/10.1098/rstb.2015.0337

Lacey, E. A., Hammond, T. T., Walsh, R. E., Bell, K. C., Edwards, S. V., Ellwood, E. R., ... & Monfils, A. K. (2017). Climate change, collections and the classroom: using big data to tackle big problems. Evolution: Education and Outreach, 10, 2, 1–13. doi.org/10.1186/s1205

McLean, B. S., Bell, K. C., Dunnum, J. L., Abrahamson, B., Colella, J. P., Deardorff, E. R., Weber, J. A., Jones, A. K., Salazar-Miralles, F., & Cook J. A. (2015) Natural history collections-based research: Progress, promise and best practices. *Journal of Mammalogy*, 97, 287-297.

Miller, G. E., Grant, P. M., Kishore, R., Steinkruger, F. J., Rowland, F. S., & Guinn, V. P. (1972) Mercury concentrations in museum specimens of tuna and swordfish. *Science*, 175, 1121-1122.

Monfils, A. K., Powers, K. E., Marshall, C. J., Martine, C. T., Smith, J. F., & Prather, L. A. (2017). Natural history collections: teaching about biodiversity across time, space and digital platforms. Southeastern Naturalist, 16 (sp10), 47–57. doi: 10.1656/058.016.0sp1008

Mooney, H., & Newton, M. P. (2012). The anatomy of a data citation: Discovery, reuse and credit. *Journal of Librarianship and Scholarly Communication*, eP1035-eP1035.

Page, L. M., MacFadden, B. J., Fortes, J. A., Soltis, P. S., & Riccardi, G. (2015) Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience*, 65, 841-842.

Pennisi, E. (2019) Report urges massive digitization of museum collections. Science, 364,115.

Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. (2019). Ecological data should not be so hard to find and reuse. Trends in Ecology & Evolution. 34:494–496. doi: 10.1016/j.tree.2019.04.005

Ponder, W. F., Carter, G. A., Flemons, P., & Chapman, R. R. (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology,* 15, 648-657.

Pyke, G. H., & Ehrlich, P. R. (2010) Biological collections and ecological/environmental research: A review, some observations and a look to the future. *Biological Reviews*, 85, 247-266.

Renaut, S., Budden, A. E., Gravel, D., Poisot, T., & Peres-Neto, P. (2018) Data management, archiving and sharing for biologists and the role of research institutions in the technology-oriented age. BioScience, 68, 400-411.

Seltmann, K., Lafia, S., Paul, D., James, S., Bloom, D., Rios, N., ... & Davis, E. (2018) Georeferencing for Research Use (GRU): An integrated geospatial training paradigm for biocollections researchers and data providers. Research Ideas and Outcomes, 4, e32449.

Shaffer, H. B., Fisher, R. N., & Davidson, C. (1998) The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution*, 13, 27-30.

Singer, R. A., Love, K. J., & Page, L. M. (2018) A survey of digitized data from U.S. fish collections in the iDigBio data aggregator. *PloS One*, 13, e0207636

Sparks, J. S., Schelly, R. C., Smith, W. L., Davis, M. P., Tchernov, D., Pieribone, V. A., & Gruber, D. F. (2014) The covert world of fish biofluorescence: A phylogenetically widespread and phenotypically variable phenomenon. *PLoS One*, 9, e83259

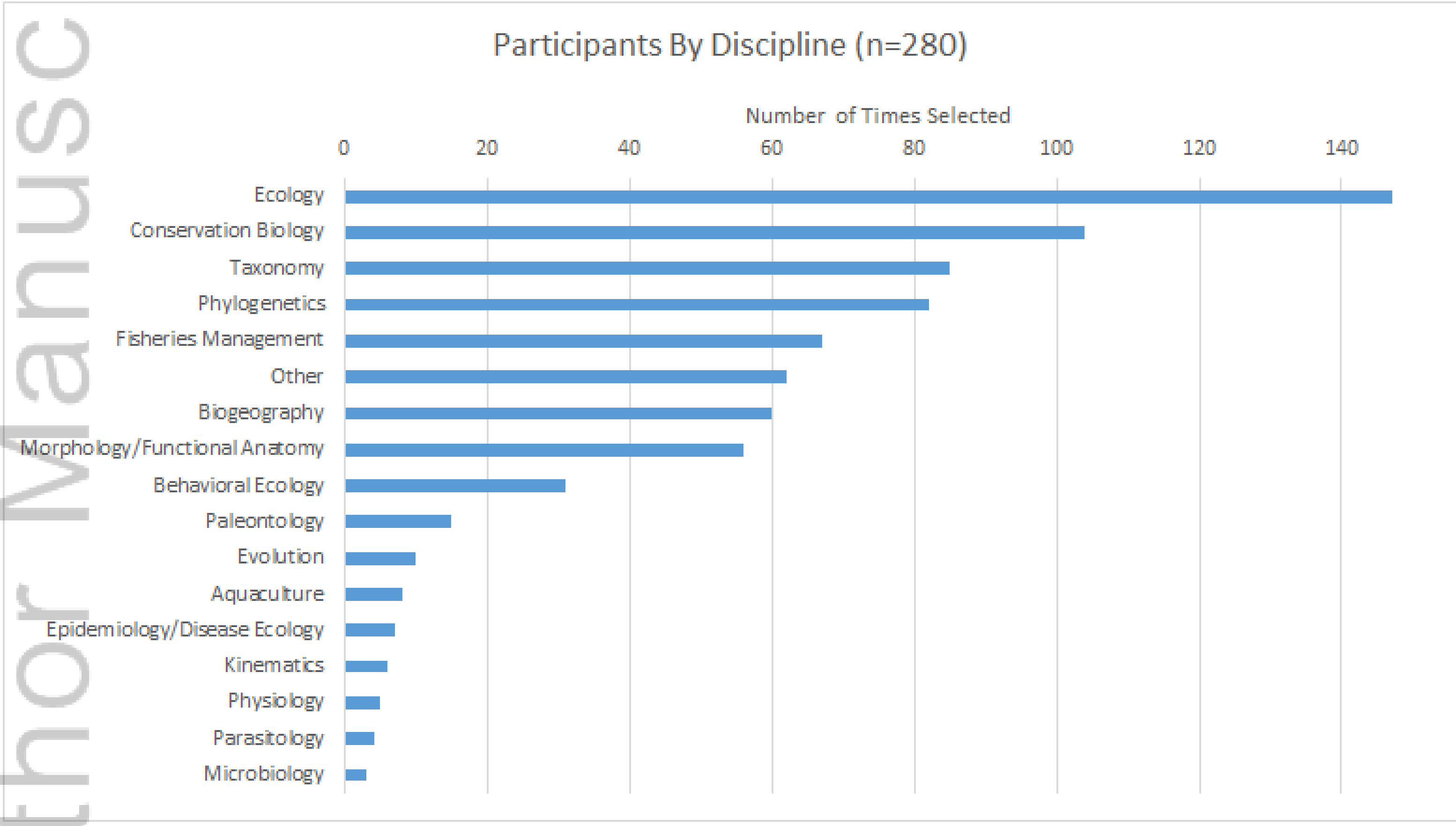
Watanabe, M.E. (2019). The evolution of natural history collections: New research tools move specimens, data to center stage., BioScience, 69 (3), 163–169. doi: 10.3897/rio.4.e32449

Winston, J. E. (1999) Describing species: practical taxonomic procedure for biologists. New York, NY: Columbia University Press.

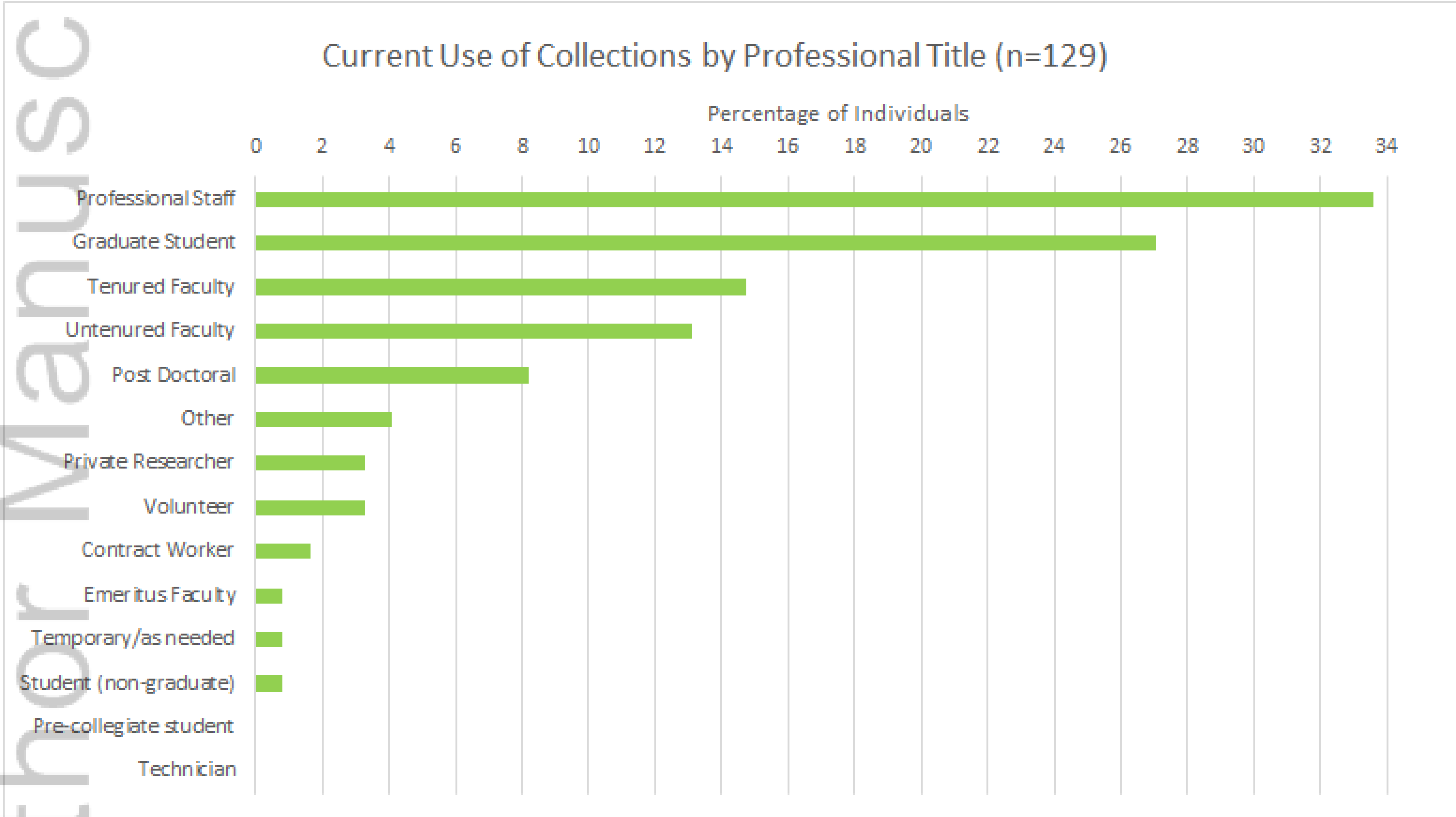
WWL (2019) Museum studies majors guide. WorldWideLearn. www.worldwidelearn.com/online-education-guide/arts-humanities/museum-studies-major.htm

SUPPORTING INFORMATION

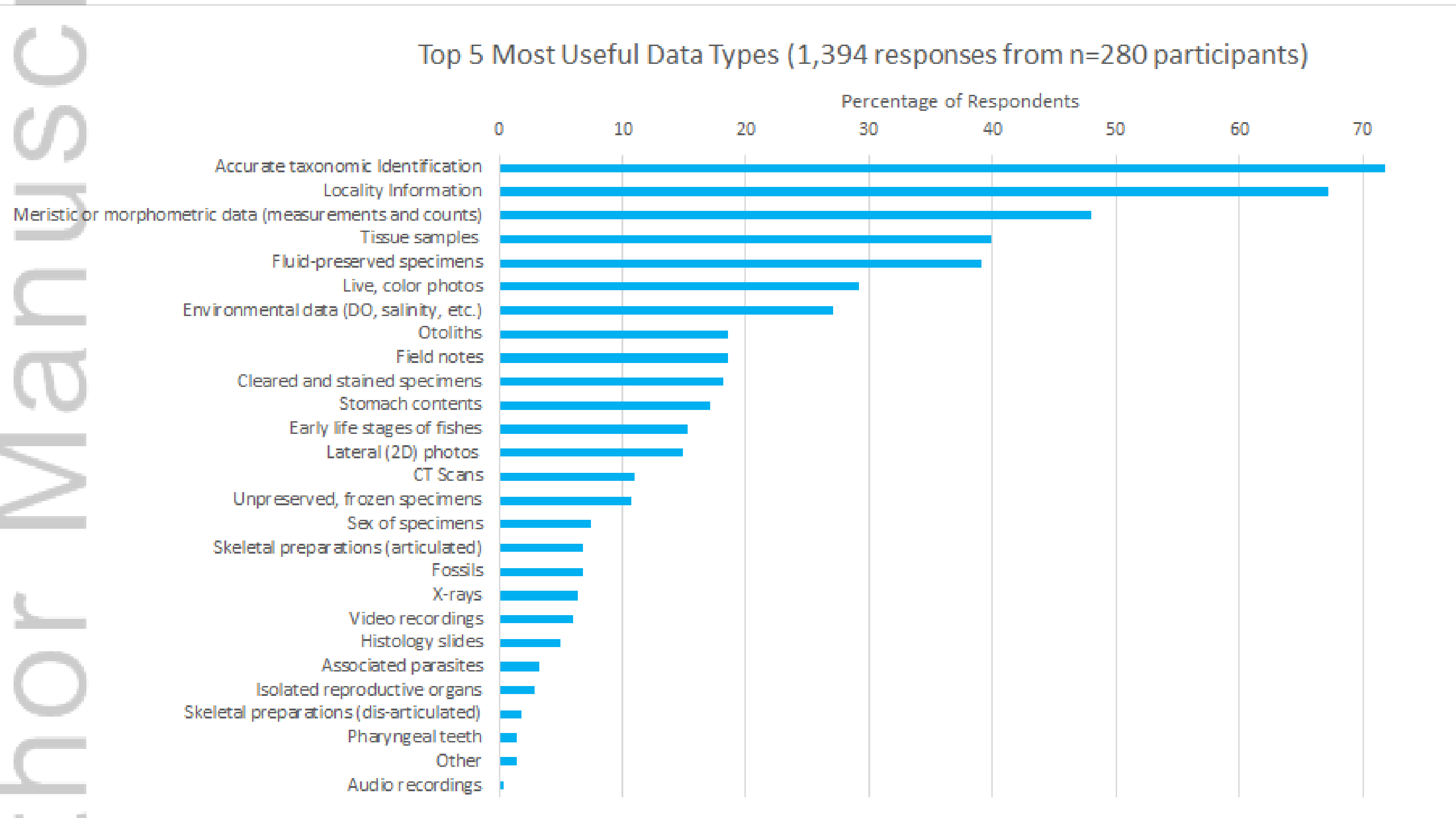
Supporting information can be found in the online version of this paper.



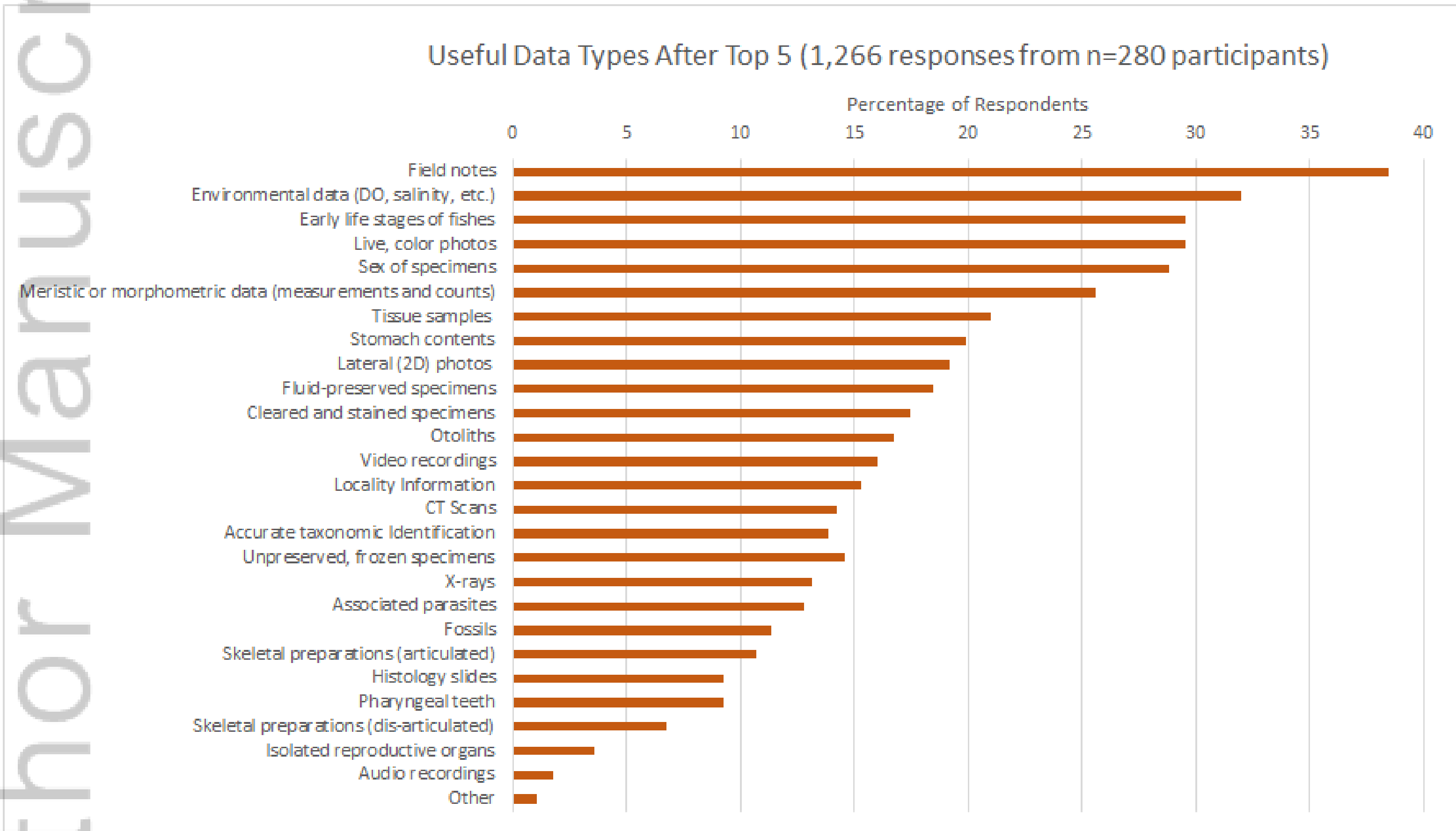
JFB_14167_Figure_1.tif



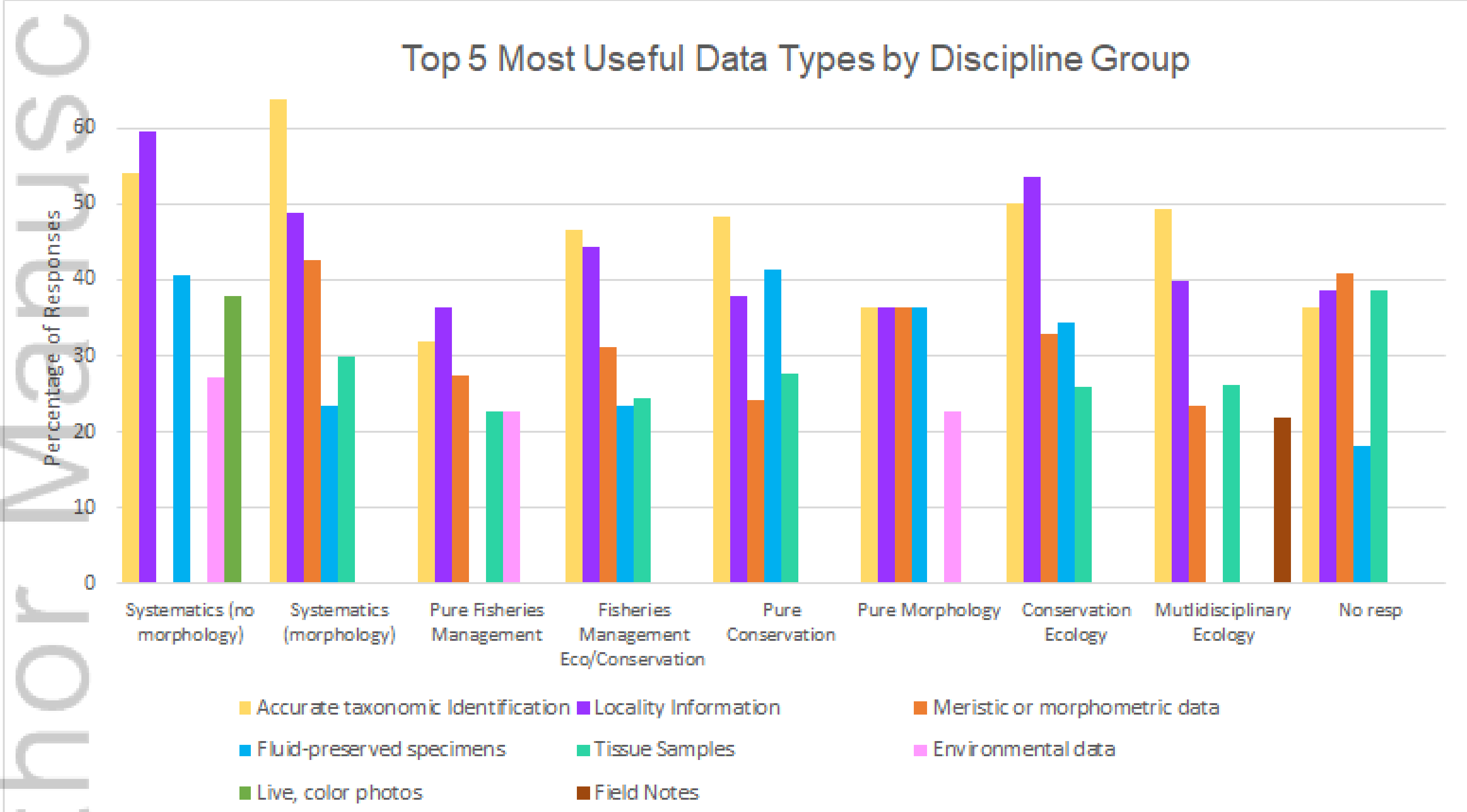
JFB_14167_Figure_2.tif



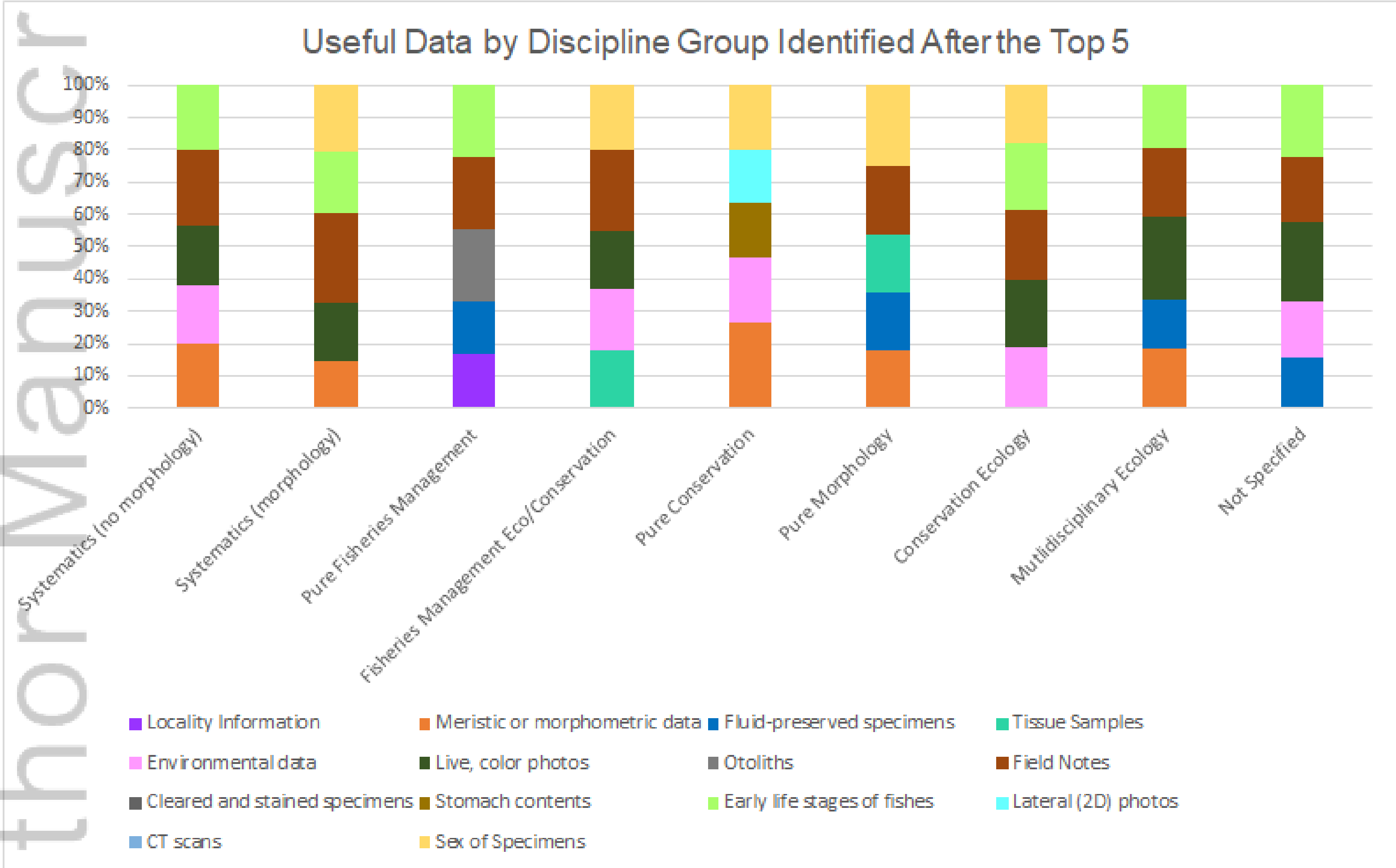
JFB_14167_Figure_3.tif



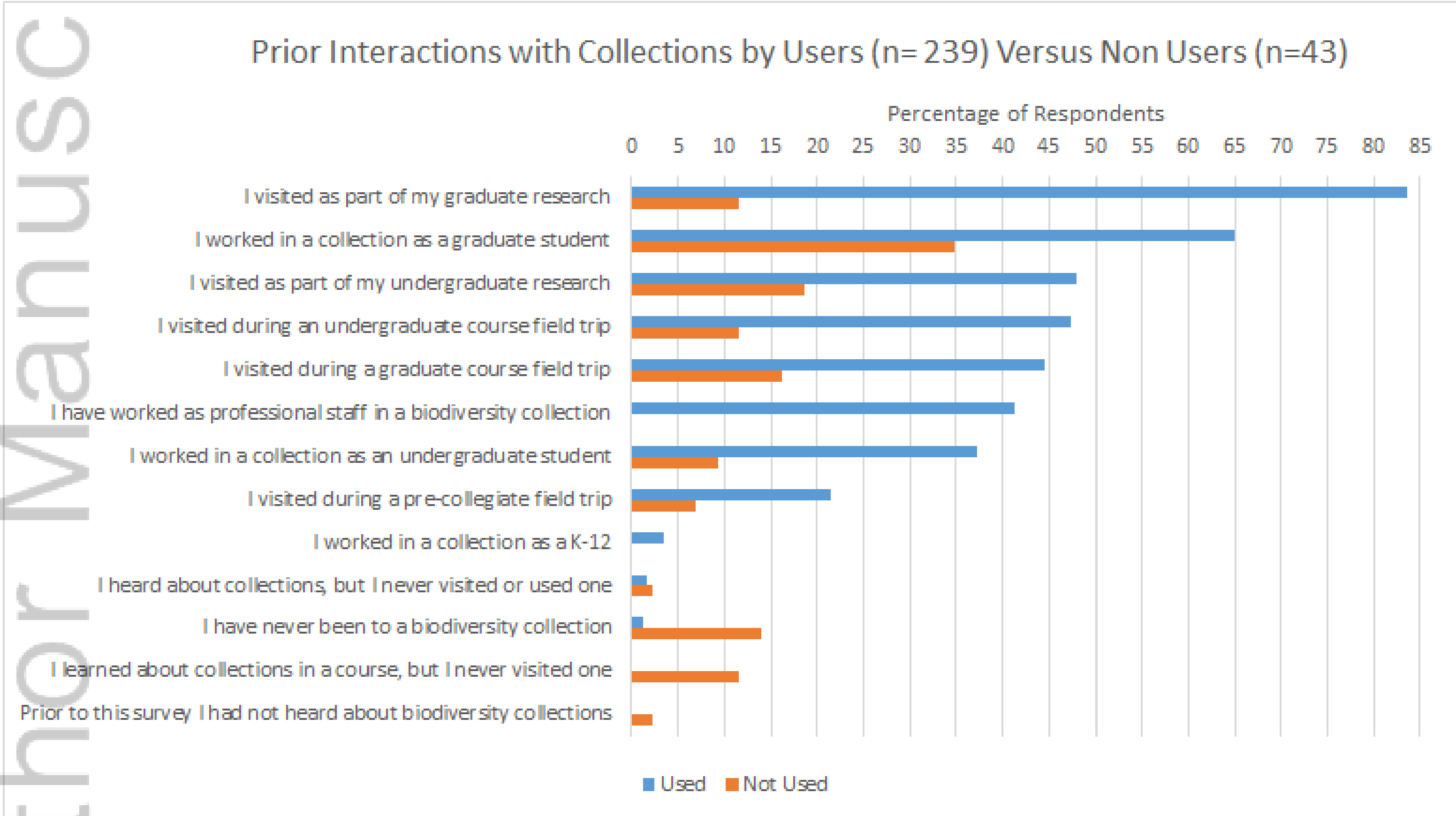
JFB_14167_Figure_4.tif



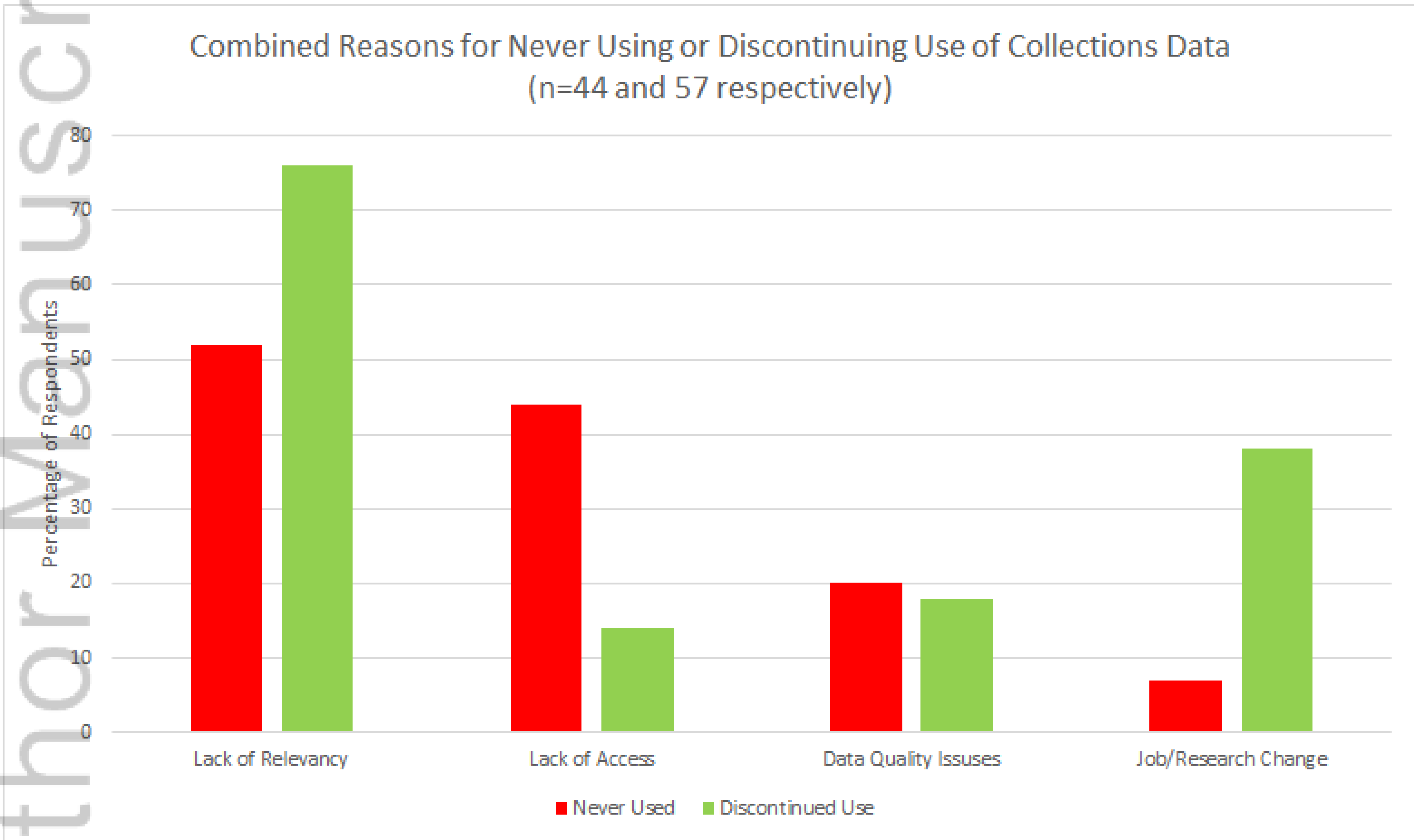
JFB_14167_Figure_5.tif



JFB_14167_Figure_6.tif



JFB_14167_Figure_7.tif



JFB_14167_Figure_8.tif

Typesetter:

1. **Change US to English spelling throughout.**
2. **Change all labels to sentence case throughout**
3. **Replace all / with an en dash (no spaces)**

FIGURE 1 Frequency distribution of survey participants' (*n* = 280) research interests by major disciplines

Typesetter: replace Participants By Discipline (n=280) Number of Times Selected with Frequency (*n*)

FIGURE 2 Frequency distribution of survey participants (*n* = 129) currently using collections by professional title

Typesetter: replace Current Use of Collections by Professional Title (n=129) Percentage of Individuals with Frequency (%)

FIGURE 3 Frequency distribution of survey participants' (*n* = 280) top five most useful data housed in fish collections

Typesetter: replace Top 5 Most Useful Data Types (1,394 responses from n=280 participants) Percentage of Respondents with Frequency (%)

FIGURE 4 Frequency distribution of survey participants' (*n* = 280) responses for data housed in fish collections marked as 'useful' after the top five selected as 'most useful'. Respondents were instructed to not select a data type they had indicated for their top 5

Typesetter: replace Useful Data Types After Top 5 (1,266 responses from n=280 participants) Percentage of Respondents with Frequency (%)

FIGURE 5 Frequency distribution of survey participants' (*n* = 280) responses for participant's top five most useful data housed in fish collections sorted by discipline group.

Typesetter

1 Delete Top 5 Most Useful Data Types by Discipline Group

2 Replace y-axis label with Frequency (%).

FIGURE 6 Cumulative frequency distribution of survey participants' (*n* = 280) responses for data housed in fish collections marked by participants after their top five most useful data were selected sorted by discipline group.

Typesetter

1 Delete Useful Data by Discipline Group Identified After the Top 5

2 Insert y-axis with Frequency (%).

3 Delete % from y-axis numerical values

FIGURE 7 Cumulative frequency distribution of survey participants identifying prior interactions with fish collections; data are sorted by user (*n* = 239) or non-user (*n* = 43) of collections

Typesetter: replace Prior Interactions with Collections by Users (n= 239) Versus Non Users (n=43) Percentage of Respondents with Frequency (%)

FIGURE 8 Cumulative frequency distribution of survey participants identifying reasons for not using (*n* = 44) or discontinuing to use (*n* = 57) collections

Typesetter

Delete

Combined Reasons for Never Using or Discontinuing Use of Collections Data
(n=44 and 57 respectively)

Replace y-axis label with Frequency (%).

TABLE 1 Survey distribution and methods for engagement in order of estimated number of individuals engaged (as of 1 July 2018).

Society, Organization or Social Media	**Estimated number of members or engagement**	**Method for survey distribution**
American Fisheries Society (AFS)	8000	Society newsletter; Listserv
American Society of Limnology & Oceanography (ASLO)	4300	Facebook
American Society of Ichthyologists and Herpetologists (ASIH)	3601	Facebook
North American Native Fishes Association (NANFA)	3129	Facebook
Facebook ichthyology page	2724	Facebook
Natural History Collections listserver (NHCOLL-L)	1920	Listserv
Desert Fishes Council	1446	Facebook
Association of Southern Biologists (ASB)	932*	Listserv
American Society of Parasitologists (ASP)	700	Listserv; Facebook
International Society of Vertebrate Morphology (ISVM)	650	Facebook
Twitter hashtag #fishsci	646	Twitter
Randy Singer Twitter page	558 (146 link clicks)	Twitter
Southeastern Estuarine Research Society (SEERS)	470*	Listserv
Southeastern Fishes Council (SFC)	204	Listserv; Facebook
Randy Singer Facebook page	57	Facebook
United States Geological Survey (USGS) – GAP analysis program	Unknown	Email

TABLE 2 Disciplines selected for each discipline group (reporting only those selected by > 50% of respondents). Percentages reflect the proportion of respondents within a discipline group that selected the corresponding discipline.

Discipline Group	Taxonomy	Biogeography	Fisheries management	Ecology	Conservation	Morphology	Phylogenetics
Systematics (no morphology; *n* = 37)	100%	84%					
Systematics (morphology; *n* = 47)	64%					100%	85%
Pure fisheries management (*n* = 22)			100%				
Fisheries management/conservation (*n* = 90)			100%	84%	66%		
Pure conservation (*n* = 29)					100%		
Pure morphology (*n* = 22)						100%	
Conservation ecology (*n* = 58)				100%	100%		
Multidisciplinary ecology (*n* = 73)				100%			