Whipple *et al. Sustainable Earth Reviews*00000000000(2024)07:130
https://doi.org/10.1186/s42055-024-00084-9

Sustainable Earth Reviews

ANALYSIS

Open Access

Leveraging virtual datasets to investigate the interplay of pollinators, protected areas, and SDG 15

Sarah Whipple[1,2*] and Stefan Moss[3]

Abstract

Biodiversity loss amplifies the need for taxonomic understanding at global, regional, and local scales. The United Nations Environmental Programme Sustainable Development Goals are explicit in their demand for greater accountability with respect to ecosystem management, and Sustainable Development Goal 15, Life on Land, specifically calls for a halt to biodiversity loss. Pollinators (bees and butterflies) are two functional groups with public attention for protection, yet little long-term data availability. National Parks, including those in the United States, act as optimal sites to study biodiversity loss, but historic data tends to vary in availability. This study addresses systematic taxonomic and digitalization biases present within historic (museum), modern (citizen science), and non-digitized (private collection) datasets for Yellowstone and Grand Teton National Parks from 1900 to 2021. We find that, although database record availability is representative of butterfly and bumble bee groups known for the area, categories such as data rescue, digitalization/availability, and management/archiving vary across database types. These findings on virtual datasets offer opportunities for conservationists to understand the efficacy of digitized collections in addressing questions of species loss over time, including the strengths and pitfalls of digitized data repositories. Additionally, virtual datasets can be utilized to monitor biodiversity under Sustainable Development Goal 15 targets while also promoting broader access to resources such as museum collections for educational purposes.

Science highlights

- Natural history collections (NHCs) work to preserve biodiversity but tend to hold taxonomic biases.
- The rapid digitalization of species occurrence data works to improve biodiversity understanding.
- Pollinator NHCs can inform conservation targets like SDG 15, but only for a subset of species.

Policy and practice recommendations

- Additional funding towards data digitalization will broaden the understanding of lesser-known taxa.
- Virtual museum resources should become more readily accessible to educate and engage the public in species conservation work; citizen science applications can act as an additional educational tool to promote public conservation interest.
- International biodiversity sampling efforts should be encouraged to document species decline.

---

*Correspondence:
Sarah Whipple
sarahemmawhipple@gmail.com
[1]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80524, USA
[2]Fort Collins Science Center, United States Geological Survey, Fort Collins, CO 80526, USA
[3]Department of Middle and Secondary Education, College of Education and Human Development, Georgia State University, Atlanta, GA 30303, USA

Keywords SDG 15, Digitalization, Pollinators, Protected areas, Biodiversity loss, Conservation, Virtual datasets, Sustainability education

Introduction

Natural history collections (NHCs) have rich histories worldwide and act as repositories for specimens of all kinds, from rare and endemic species of flowers to pests and invasive insects. However, the potential of NHCs to explore landscape changes and project species conservation targets like those set by the United Nations Environmental Programme (UNEP) Sustainable Development Goals (SDGs) 14 (Life Below Water) and 15 (Life on Land) is largely overlooked [1, 2]. NHCs have been used to highlight the impacts of climate change on biodiversity, especially for insects and plants, but these efforts for pollinator species have often been restricted to a few charismatic groups such as the Nymphalidae and Lycaenidae butterflies, small regions, and short time periods [3]. Researchers have yet to connect biodiversity interactions over time back to species' phenological changes, and NHCs provide opportunities to execute such connections [4].

As highly mobile and diverse species, insects have historically been understudied and undervalued, making them more challenging to conserve [5, 6]. The Endangered Species Act (ESA) of the United States (US) includes 97 federally listed insect species; in contrast, 419 vertebrate species are considered threatened or endangered [7]. The difficulties accredited to insect protection under the ESA include their dynamic relationships with plants that may also have listing status, habitat-specific requirements, and limitations in taxonomic understanding [6, 7]. In some states within the US, insects are not considered wildlife and, as a result, cannot be conserved under laws like the ESA [8]. When considered at a broader, global scale, biodiversity conservation efforts for taxonomic groups such as insects are currently prioritized through Sustainable Development Goal (SDG) 15—Life on Land [9]. The United Nations established seventeen SDGs in 2012 to address the economic, environmental, and political challenges facing countries all around the world. SDG 15 contains twelve targets and fourteen indicators, which together serve as a roadmap for addressing issues related to forest management, land use and species extinction. SDG 15 aims to "protect, restore, and promote sustainable use of terrestrial ecosystems…and halt biodiversity loss" [10]. Biodiversity loss and climate change impact critical ecosystem services such as pollination, an interaction between plants and insects that produces over 70% of global food sources worldwide, which also intersects other SDGs beyond SDG 15 such as SDG 2—Zero Hunger [9, 11]. The relationships and interactions between SDGs are still being studied, but evidence suggests that although it may be impossible to achieve all of them, it is still important to consider how they influence each other in terms of synergistic benefits and trade-offs [12]. Almost 20% of the most common insect pollinator "flower visitor" functional groups—bees and butterflies—face extinction threats within the next generation; however, the patterns and causes of decline are still in question [11, 13].

The historical reference of pollinators (bees and butterflies) is of particular concern as species decline occurs nationally and internationally [13]—because of this, there is an "urgency" to document species diversity through NHCs and other data repositories [14]. The inability to determine the patterns and causes of decline underscores the need for efforts such as digitalization. NHC digitalization efforts are being accelerated to document species facing biodiversity loss through several initiatives that integrate biological collections into education and research projects [15] with available taxonomic and genetic information. The Symbiota Collection of Arthropod Network (SCAN) acts as the primary NHC digitalization repository for North American arthropod collections (185 collections in total), with other large-scale biodiversity collection tools such as the Global Biodiversity Information Facility (GBIF) and iDigBio acting as larger repositories for all taxa within NHC digitalization discussions [14]. Further emphasizing the species documentation "urgency," in an initial synthesis of North American collections, Cobb et al. [14] found that ~95% of North American insect specimen labels have yet to be transcribed for research purposes, and about two percent of specimens have been digitized with images (see Additional File 2 for digitalization statistics relevant to this study). With US NHC target digitalization trajectories, the group projects that 38% of all current North American arthropod specimens can be digitized by 2050, and less than one percent of collections will be digitized with images.

More recent introductions of large-scale datasets through citizen science, where the public contributes toward scientific efforts [16], have the potential to elucidate species patterns of decline [17], in addition to the use of NHCs and other research efforts. Emerging mobile phone technology has increased the collection and distribution of citizen science datasets, and with improvements to taxonomic and locational accuracy since their inception [18]. By complementing data sources from taxonomic experts and community scientists, species assessments can be done during the age of NHC database digitization [15] and with the assistance of increased public participation in the scientific and biodiversity collection processes [16]. For the purposes of this analysis, we consider NHCs and citizen science databases such as iNaturalist and BugGuide as synonymous resources that can help researchers ascertain species occurrence patterns over time. Citizen science datasets will continue to become richer in both data quality and quantity as technology improves and public interest in scientific discovery grows [19]. However, we cannot dismiss that citizen science data does present some limitations, such as the types of questions that can be asked by a researcher [20], the effort (and funding) required to successfully implement community help and long-term engagement [21], data biases, and data accuracy [22, 23]. Nevertheless, these limitations are outweighed by the long-term benefits of such data collection and its ability to provide understanding to current data gaps within geographic regions.

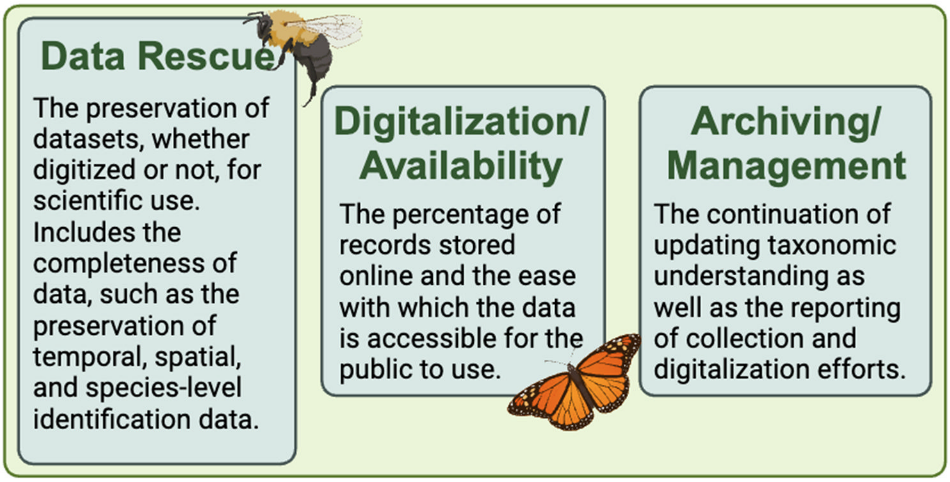
As NHC records become available to the public through the digitalization of specimens, these data will become useful in conjunction with citizen science and other recent biodiversity species inventory collection efforts, and at the global scale [1, 3]. Recent studies have also noted the potential of digitalization as a powerful tool in achieving the SDGs centered around climate change, species conservation, as well as different sectors such as agriculture, energy, and health [24, 25]. However, digitalization capabilities vary across collections and geographic regions, with percentages of complete databases tied to the size of the collection, resources available, and funding tied to taxa-specific questions [14]. Taxonomic and geographic biases present within species occurrence datasets challenges researchers, land managers, and policymakers alike in their capabilities to effectively work towards SDG 15, because a baseline understanding of species diversity is critical before significant conservation actions can occur [26]. However, with the addition of citizen science data repositories that act as real-time data collection tools, researchers can assess species occurrences and upload observations more rapidly than the digitalization process allows for, and this can help in the understanding of species changes documented globally over time [27].

When connected with multiple records, species groups, and field sampling protocols, NHC studies have successfully tracked progressions in insect flight periods, temperature responses, genetic variation, and voltinism stages that are all tied to warmer temperatures and other climatic pressures, but these efforts have not happened within the Rocky Mountain region of the US, or more specifically, the Greater Yellowstone Ecosystem (GYE) and within Yellowstone and Grand Teton National Parks [3, 28]. US National Parks, in addition to other protected areas nationally and worldwide, can act as species refuges for all biodiversity, including pollinators [5]. In the age of these climatic pressures, prioritizing suitable habitat for species conservation is essential, as addressed through SDG 15 [29]. Both Yellowstone and Grand Teton National Parks have historical records in federal, academic, and local museums, as well as private collections, but there are less known collection efforts in the neighboring national forests and private lands within the GYE, one of the largest contiguous protected areas in the continental US, and no synthesis of all available historical records in virtual databases. This is a common problem observed across biodiversity studies intending to utilize digitized records, as the compilation of records across differing resources takes time, effort, curatorial support, and taxonomic expertise [30]. As such, this study aims to act as a benchmark for understanding US pollinator decline within two protected areas (Yellowstone and Grand Teton National Parks) as told through historic collections.

For this study, “collections” are defined as those that are publicly available through online databases. This includes resources coming from large biodiversity repositories, such as iDigBio, GBIF, and SCAN, smaller biodiversity repositories, such as individual university databases, citizen science databases such as iNaturalist, and a review of the National Park Service (NPS) research permitting reports relevant to bumble bees and butterflies. When a non-digitized dataset was known, such as a private collection or a count list, these resources were also included. Incorporating both citizen science records and permit-reported data is a method aimed at addressing uncertainties associated with digitized data. Additionally, it acknowledges the significance of records that may not fit within the conventional museum curation framework but are nonetheless valuable resources due to public participation. This effort also aims to rank the status of species diversity records within the known databases in their ability to answer questions surrounding data rescue, digitalization/availability, and archiving/management, which translates beyond the parks to larger SDG 15 and sustainability education targets surrounding biodiversity conservation efforts and systematic digitalization priorities (Fig. 1).

Study goals

This study offers a comprehensive overview of pollinator (bumble bees and butterflies) data within Yellowstone (YELL) and Grand Teton (GRTE) National Parks. We aim to highlight past collection priorities and shed light on the status of species biodiversity through these collections. The following questions guided this research: what pollinator species (bumble bees and butterflies) were known or documented to occur within YELL and GRTE from 1900 to 2021, and what patterns were



Data Rescue

The preservation of datasets, whether digitized or not, for scientific use. Includes the completeness of data, such as the preservation of temporal, spatial, and species-level identification data.

Digitalization/ Availability

The percentage of records stored online and the ease with which the data is accessible for the public to use.

Archiving/ Management

The continuation of updating taxonomic understanding as well as the reporting of collection and digitalization efforts.

Fig. 1 Working definitions for data rescue, digitalization/availability, and archiving/management used when assessing species occurrence databases. Figure created with Biorender.com

observed within virtual databases regarding digitization and species understanding efforts? Broadly, pollinator data for groups such as bees and butterflies are more available in collections because of their charismatic species groups [14] and benefits to ecosystem services [13]. However, taxonomic and database digitalization biases present within current datasets are uncertain, and therefore, data for certain species within these groups may be limited. Before addressing questions such as species range shifts or pollinator declines, baseline data understanding is needed that can act as evidence for any species changes that have occurred. This is especially important as ecosystems face panarchy, or irreversible rates of change, and species will need to become more resilient towards shifting landscapes or risk rapid rates of decline [31–35]. NHCs and other virtual datasets, as a result, may act as critical time capsules for biodiversity by documenting resilient and non-resilient species in perpetuity. This information can then be utilized to formulate effective conservation strategies and contribute to achieving SDG 15 targets.

Through this study, we evaluated the availability of collection records compared to expected species occurrence records (data robustness analysis), analyzed the biodiversity and evenness of such records (species diversity analysis), and compared databases with records used in this study for taxonomic and digitalization gaps (database gap analysis). We hypothesize that the current knowledge of pollinators is biased towards certain taxonomic groups, leading to gaps in species understanding and potential underestimations of the importance and conservation needs of other pollinator groups. Taxa that are more charismatic and showier, such as the Nymphalidae butterflies and bumble bees, will be more prominent in the databases than lesser-known and smaller taxa, such as the Hesperiidae butterflies. Similarly, databases will be biased in the data reported and available for varying taxonomic groups and spatial areas of the parks, even among the bumble bee and butterfly groups. These data digitalization biases will reflect not only the collector histories stored within the NHCs but also the resources, personnel, and taxonomic specialties of the subsequent collections. Taxonomic and database digitalization biases, therefore, may significantly influence one's ability to inform overarching questions such as biodiversity loss for broad groups such as pollinator species in western US National Parks, and this could have broader impact on targets such as SDG 15 in its intentions to conserve species diversity.

Methods

Dataset preparation

All digitized NHC and citizen science records available for the counties within YELL and GRTE National Parks and surrounding areas from 1900 to 2021 were tracked for two functional groups of pollinators: butterflies (order Lepidoptera) and bumble bees (order Hymenoptera, family Apidae, genus *Bombus*). Online species NHC downloads were filtered first by location, a 60-km bounding box containing the two parks and surrounding GYE area, then by the respective taxonomic level. Within butterflies, species were filtered based on the five butterfly families observed within the park: Hesperiidae (Skippers), Lycaenidae (Blues and Hairstreaks), Nymphalidae (Brush-footed Butterflies), Papilionidae (Swallowtails), and Pieridae (Whites, Sulphurs, and Yellows). Datasets with relevant, digitized records included: GBIF [36–41], SCAN, the Smithsonian Institution, *Bombus* of Canada, the Lepidopterist's Society, iDigBio, BugGuide, and iNaturalist. These databases were selected as primary online resources given their frequency of use within the NHC and entomology literature [14]. For a full list of repositories housed within databases such as GBIF, SCAN, and iDigBio, including federal, state, and university repositories, see Additional Files 1 and 2.

Datasets with known relevant, non-digitized records included: Yellowstone's Fourth of July Butterfly Count records (Marilyn Lutz, NPS, Joshua Tree, California, personal communications, September 30, 2018), the Yellowstone Heritage and Research Center, and the Harp Collections (Chuck Harp, Colorado State University, Fort Collins, Colorado, personal communications, January 20, 2019). To account for gaps in species data that have yet to be digitized but may be stored within NHCs or private collections, the NPS research permit and reporting system database was searched for both GRTE and YELL by reviewing research investigator annual reports (IAR) publicly reported since 1991 with ties to pollinator, bumble bee, and butterfly research. The IAR system within the NPS requires data reporting prior to permit resubmission, so this database can provide baseline data from each approved research project done in the park, even if a project did not collect specimens or have the resources to digitize their specimen collections. The database was searched by each park with the key terms "pollinator(s)," "bees," "*Bombus,*" "butterfly(s)," and the five butterfly families individually; this ensured all relevant permits were viewed. Data that included species-level identifications and specimen counts were added to the list of available historic records.

When available, data from online repositories were queried using the DarwinCore format, a biodiversity archive standard that includes taxon, occurrence, and event metadata [42]. This ensured that duplicates present within overlapping databases, specifically within data repositories such as GBIF, iDigBio, and SCAN, could be filtered out of the final analysis. All data were prepped using the "tidyverse" package in R [43]. Databases were cleaned based on their robustness of records; first, by the number of total occurrences and its proportion of digitized (i.e., records with complete metadata, including images) records, as well as records with complete taxonomic, georeferenced, and temporal information. Records that were incomplete were flagged but not omitted from the final analysis. All data, including the analysis, are available for download and use on Mendeley Data [44].

Virtual database robustness analysis

We followed a three-step analysis of available data to ensure that comparisons between species diversity and database completeness could effectively occur with the data available in a digitized format. For an overview schematic of this analysis process, including the data components and research aims, see Fig. 2. First, we evaluated the online databases (including the NPS research permits) and in-person collections for data quality and quantity. To answer this question, we ran a chi-square (χ^2) test in Microsoft Excel to compare the relationship between butterfly species families and genus *Bombus* observations within the databases ($n = 47$ databases). We organized our χ^2 test by using a list of expected species observed in the park and compared this to the observed list of species available across the databases. *P*-values less than 0.05 indicated that database records were significant in comparison to expected species occurrences. If the observed list of species was not representative of what was expected, we would have needed to restructure the rest of the analyses to reflect these data gaps apparent at the onset of this effort.

Species diversity data analysis

Next, we calculated species richness and evenness indices across both parks, outside of the parks, and in the overall area using all online, in-person, and research permit records with species-level identifications. For this, the Shannon-Wiener Index (H'), Pielou evenness (J), and Shannon-Wiener Effective Diversity Number ($e^{H'}$) indices were used [45, 46]. Shannon-Wiener Index, Pielou evenness, and effective (true) Shannon-Wiener Index diversity values were calculated across taxonomic groups. All calculations were performed

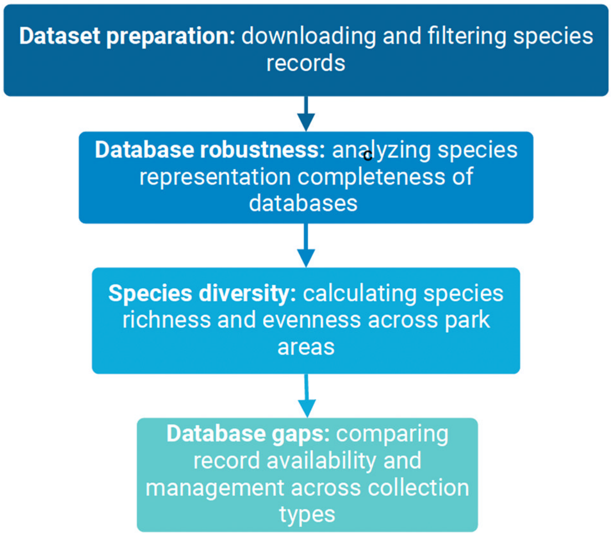


Fig. 2 The methods workflow used to prepare and analyze data for species occurrence understanding and taxonomic/database gaps. Figure created with Biorender.com

using the "vegan" package in R [47]. Higher Shannon-Wiener Index values represent more diverse areas of species diversity; these values typically fall within the range of 1.5–3.5. Pielou evenness values closer to one indicate a richer, more evenly dispersed species diversity seen across the area. Like the Shannon-Wiener Index, higher Shannon-Wiener Effective Diversity Number indices represent true, diverse equivalent areas of species diversity that are minimized by sampling biases such as larger abundances of species within the sample size compared to rare species [48]. Because of the diversity and sampling uncertainties presented within the Shannon-Wiener and Pielou evenness indices, the Shannon-Wiener Effective Diversity Number was used to provide an equivalent, hypothetical understanding of species diversity to the observed ecosystem by balancing out rare versus abundant species. We did this to minimize sampling biases seen across taxa, areas, and years within the NHCs and citizen science databases, as this is something that is criticized within richness and evenness studies that utilize these analyses for biodiversity understanding [48].

Database gaps analysis

Finally, given gaps in data availability within online databases due to the digitalization process, we ranked all occurrence data in relation to categories of data rescue, digitalization/availability, and archiving/management. Each database category (federal, citizen science, state, private, mixed-source, and university records) was scored using a Spearman rank correlation method in Microsoft Excel based on digitalization metrics. Digitalization metrics relied on database reporting available online and followed a 0–100 scale for the data rescue, digitalization/availability, and archiving/management categories. A Spearman rank correlation coefficient less than the critical value for n = 6 indicated a significant correlation between categories and databases.

Results

For an overview understanding of available records by database type and taxonomic group, as well as a database overview and digitalization statistics, see Additional Files 1 and 2. The following records could be refined based on currently known ranges: *Bombus lapponincus*, four records; *B. terricola*, one record; and *B. vosnesenskii*, two records. For butterflies, the following species only had one observation within the databases, or species identifications could be refined based on currently known ranges: (*Anthocharis cethura, Euchloe lotta, Pieris oleracea, Pieris virginiensis, Papilio canadensis, Oeneis alberta, Oeneis macounii, Euphilotes glaucon, Cupido comyntas, Satyrium acadica, Erynnis pacuvius lilius, Hesperia leonardus, Hesperia ottoe, Hesperopsis alpheus, Megathymus streckeri, Oarisma edwardsii, Hesperopsis libya, Polites rhesus, Polites vibex,* and *Pompeius verna*). Eleven records were identified only to the family level (four Lycaenidae, two Pieridae, and five Rionidae (Metalmarks)), and three records within the Nymphalidae were only identified at the subfamily level (Limenitidinae). Within *Bombus*, 39 records from online databases were only identified down to the genus level, and 213 records had no locational or temporal information available.

Virtual database robustness

To address database robustness in species diversity understanding, χ^2 values less than 0.05 at the 95% confidence interval indicated there was a statistically significant relationship between species observed across databases. χ^2 results show that the genus *Bombus* and all butterfly families have significant database representation for the observed species occurrences compared to the expected values ($p < 0.05$) (Table 1). The total collections available for each taxonomic group varied by location (n = 10,051 records), with *Bombus,* Lycaenidae, and Nymphalidae groups having the largest digitalization of records. For a breakdown of the most common species collected, see Fig. 3.

Species diversity

In GRTE, YELL, outside of the parks, and within the GYE overall, family Nymphalidae had the highest Shannon-Wiener (*H'*) (3.34, 3.225, 2.92, and 3.38, respectively) and Effective Shannon-Wiener Index ($e^{H'}$) (28.21, 25,70, 18.53, and 29.33, respectively), while family Papilionidae had the lowest values across all locations (0.59 H' for GRTE and YELL and 0.83 for GYE; 1.80 $e^{H'}$ for GRTE and YELL and 2.30 for GYE) except for outside of the parks, which had the lowest value for genus *Bombus* (1.66 *H'* and 5.27 $e^{H'}$, respectively) (Table 2). For Pielou evenness (*J*), family Pieridae had the highest value

Table 1 χ^2 values for each taxonomic group observed across the databases with GYE records (n = 47), and *p*-values to represent database significance in representing expected species

	Bombus	**Hesperiidae**	**Lycaenidae**	**Nymphalidae**	**Papilionidae**	**Pieridae**
χ^2	5022.05	1985.43	1283.85	2405.66	1086.32	435.22
p-value	0*	0*	8.82×10^{-239}*	0*	1.76×10^{-197}*	8.27×10^{-65}*

P-values less than 0.05 are denoted with an asterisk (*), meaning that the observed database representation is significant, and representative of the population compared to the expected values

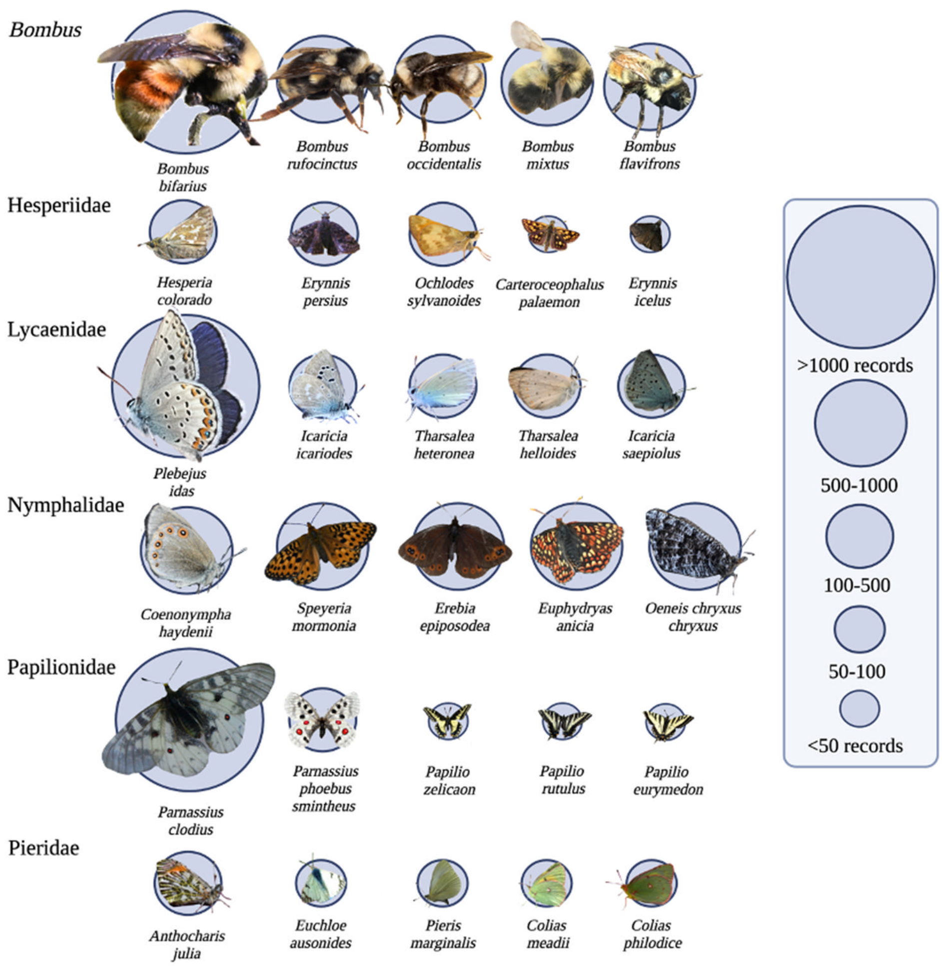


Fig. 3 Species breakdown of most common occurrences by taxonomic group. For genus *Bombus, B. bifarius* = 1643 records, *B. rufocinctus* = 902 records, *B. occidentalis* = 765 records, *B. mixtus* = 560 records, and *B. flavifrons* = 537 records. For family Hesperiidae, *Hesperia colorado* = 105 records, *Erynnis persius* = 57 records, *Ochlodes sylvanoides* = 54 records, *Carterocephalus palaemon* = 42 records, and *Erynnis icelus* = 29 records. For family Lycaenidae, *Plebejus idas* = 1820 records, *Icaricia icariodes* = 259 records, *Tharsalea heteronea* = 198 records, *Tharsalea helloides* = 137 records, and *Icaricia saepiolus* = 127 records. For family Nymphalidae, *Ceoenonymha haydenii* = 418 records, *Speyeria mormonia* = 363 records, *Erebia epiposodea* = 183 records, *Euphydryas anicia* = 181 records, and *Oeneis chryxus chryxus* = 180 records. For family Papilionidae, *Parnassius clodius* = 1299 records, *Parnassius phoebus smintheus* = 427 records, *Papilio zelicaon* = 51 records, *Papilio rutulus* = 23 records, and *Paplilio eurymedon* = 18 records. For family Pieridae, *Anthocharis julia* = 118 records, *Euchloe ausonides* = 88 records, *Pieris marginalis* = 76 records, *Colias meadii* = 72 records, and *Colias philodice* = 70 records. All photograph credits of S. Whipple, G. Bowser, and additional field interns

in GRTE (0.32), YELL (0.32), and across the overall area (0.31). Family Hesperiidae had the highest evenness value for the outside of the park records (0.45). In contrast, the genus *Bombus* had the lowest evenness value outside of park sampling areas (0.25), the Papilionidae had the lowest value in GRTE (0.13) and YELL (0.13), and the Lycaenidae had the lowest overall value (0.20). More variation occurred across families for middle ranking richness and evenness values. There were minimal historic or citizen science records available for Papilionidae outside of the parks (five records), which caused these values to be omitted from the species richness and evenness rankings.

Database gaps

In the analysis, a Spearman rank correlation coefficient below the established critical value of 0.829 for n = 6 was indicative of a significant correlation between categories and databases. The Spearman rank correlation results indicate that the relationship between data rescue and archiving/management, as well as digitalization/availability and archiving/management was significant across the database categories (above the Spearman rank critical value of 0.829 for 0.05 significance) whereas the relationship between data rescue and digitalization/availability was not significant (below the Spearman rank critical value of 0.829) (Table 3). When ranking across

Table 2 Shannon-Wiener Index (*H'*), Pielou evenness (*J*), and effective (true) Shannon-Wiener Index diversity ($e^{H'}$) values for specimens across families collected in GRTE, YELL, outside of the parks, and GYE-wide

	Bombus	**Hesperiidae**	**Lycaenidae**	**Nymphalidae**	**Papilionidae**	**Pieridae**
Shannon-Wiener Index (H')						
GRTE *H'*	2.14	2.33	2.26	3.34***	0.59*	2.30
YELL *H'*	2.35	2.77	1.79	3.25***	0.59*	2.47
Outside of parks *H'*	1.66*	1.70	2.29	2.92***	0#	1.91
GYE *H'*	2.29	2.65	2.09	3.38***	0.83*	2.49
Effective Shannon-Wiener Index diversity ($e^{H'}$)						
GRTE $e^{H'}$	8.53	10.24	9.55	28.21***	1.80*	10.00
YELL $e^{H'}$	10.51	15.94	5.97	25.70***	1.80*	11.84
Outside of parks $e^{H'}$	5.27*	5.45	9.91	18.53***	1#	6.72
GYE $e^{H'}$	9.87	14.11	8.12	29.33***	2.30*	12.02
Pielou evenness (J)						
GRTE *J*	0.26	0.31	0.22	0.24	0.13*	0.32***
YELL *J*	0.28	0.30	0.19	0.24	0.13*	0.32***
Outside of parks *J*	0.25*	0.45***	0.37	0.30	NA	0.43
GYE *J*	0.27	0.30	0.20*	0.24	0.20	0.31***

Shannon-Wiener values that are larger represent more diverse areas based on species diversity. Pielou evenness values that are closer to one represent more evenly diverse areas based on species diversity. Asterisks are used to represent highest values (***) compared to lowest values (*) across taxonomic groups and areas. (#) indicates the presence of no historic or citizen science records for the area

the three categories, database categories such as citizen science repositories scored highly in their data rescue, digitalization efforts, and management. Out of all categories, private, non-digitized resources scored the lowest in their rankings across all three categories.

Table 3 Database category rankings for the three database review categories: data rescue, digitalization/availability, and archiving/management

Rank correlation			
	Data rescue	**Digitalization/ availability**	**Archiving/ management**
Citizen science	95	100	75
Federal	75	5	50
Mixed	75	85	75
Private resources	5	0	0
State repository	25	25	25
University	50	25	50
	Data rescue rank	**Digitalization rank**	**Archiving rank**
Citizen science	1	1	1.5
Federal	2.5	5	3.5
Mixed	2.5	2	1.5
Private resources	6	6	6
State repository	5	3.5	5
University	4	3.5	3.5
	Data rescue + digitalization	Data rescue + archiving	Digitalization + archiving
Spearman correlation	0.735294118	0.925476223	0.85084104
N = 6			
Critical value for *p*-value < 0.05	0.829		
	Not significant	Significant**	Significant**

Initial rankings followed a 0–100 scale and were generated based on digitalization statistics available online. Critical values higher than 0.829 at the 95% significance level indicated a correlation between categories

Discussion

This study examined pollinator species within Yellowstone and Grand Teton National Parks from 1900 to 2021 and highlighted biases in digitalized databases towards more charismatic taxa like Nymphalidae butterflies and bumble bee species. The data indicate that previous pollinator sampling efforts within Yellowstone and Grand Teton National Parks varied across taxa, collections, and regions. In addition, the transition of data from NHCs to citizen science applications could be an effect of museum digitization backlogs or due to the rising community present on citizen science platforms. Private collections were informative for some species groups (Hesperiidae) that lacked data among other platforms and locations; however, the trade-offs between database types limited the efficacy of species comparisons amongst datasets. This is potentially leading to an underestimations of other pollinator groups' importance and conservation needs, and thus affecting broader biodiversity conservation goals such as protection under the ESA or SDG 15. NHCs, citizen science applications, and private/permit-derived data, nonetheless, highlighted a handful of taxa that researchers can use as occurrence baselines for future species conservation monitoring, despite the taxonomic and database biases present.

The Chi-Square test provided baseline evidence of the database's completeness that could inform subsequent patterns of species richness and evenness within known, digitized NHC and citizen science records. This step was critical given the uncertainties in data availability based on variations in digitalization efforts across repositories. The gaps we found in the study are that the Shannon-Wiener Index, Pielou evenness, and effective Shannon-Wiener Index diversity indices emphasize the Nymphalidae species richness seen within the GYE and on a global scale (Table 2) [49]. While family Papilionidae is not the richest or even in species diversity in the GYE, these trends correlate to this group having the fewest number of species within a butterfly family observed in the GYE [50] and to the previous researchers who had taxonomic biases towards the *Parnassius* butterflies in GRTE and not necessarily within YELL (Fig. 3) [51, 52]. Outside of the parks, private collectors with interests in the family Hesperiidae led to higher observations of this group and the species diversity present (rather than a taxonomic focus on a particular species), and this caused the higher species evenness value observed (Chuck Harp, Colorado State University, Fort Collins, CO, personal communications, January 20, 2019). Lastly, *Bombus* richness and evenness values align with the taxonomic structure of the order Hymenoptera, family Apidae [53]; for the GYE, there are other groups within Apidae that are more robust in species diversity but lack digitized record availability. As the digitalization of more repositories occurs, these values of richness and evenness will change to be more consistent with the weight of insect species groups across broader taxonomic scales.

However, as the digitalization of NHCs evolves over time through the increased pervasiveness of improving technology and the incorporation of tools such as automated content recognitions and artificial intelligence (AI), these metrics will certainly change, as will the systematic taxonomic biases observed (Fig. 3); these taxonomic biases, as predicted, align with the assumption that collection biases occur towards charismatic, well-known species rather than rare, small, or lesser-known taxa [26, 54]. There are species for which the understanding of its range has changed, or additional time could be devoted to improving a collection's taxonomic understanding. The research permitting review identified 24 unique projects and 116 permits in total, many of which captured specimens from the park for identification or genetic purposes. These efforts provide targeted understanding of certain species but leave data gaps in the status of curation needs; this is a common problem seen in entomological studies, where taxonomic expertise requirements for specimen collections lead to delays in data processing [55]. High-level assessments continue to lump biodiversity loss into species groups, and as a result, such conclusions may be inaccurate or incomplete [56]. The global decline of entomofauna, with 40% of all insect species facing extinction due to habitat loss, accelerates the need to more adequately assess their status using digitalization [5].

Data gaps have been noted as key issues in achieving the SDGs [57]. A robust understanding of species and their unique conservation needs is critical to SDG 15 considering the important ecosystem services that biodiversity provides. By filling data gaps, information becomes more reliable and coherent, which leads to better decision making regarding SDG 15 conservation initiatives [58, 59]. Similarly, biased data fail to provide an adequate assessment of trends in biodiversity loss, and the inability to capture the full scope of species impact has implications for their long term sustainability [60]. Species digitalization can improve these outcomes by removing constraints, and making data more readily accessible to the general public for analysis and interpretation. Although digitalization can be a powerful tool for great efficiency and innovation for achieving SDG 15, its use is not without potential negative consequences such as energy intensive mining and e-waste production [61]. Increased use of technology and digital systems promotes economic growth, and results in changes in human behavior that may adversely affect other SDGs such as SDG 12, responsible consumption and production, and SDG 8, decent work and economic growth. Research related to the relationships between SDGs is ongoing and suggests a high level of complexity with the need to balance potential benefits and trade-offs [12]. Since digitalization is viewed as cross-cutting, with potential synergistic impacts, its positive effect on SDG 15 should be viewed in context with other SDGs [62].

Additionally, the Spearman rank correlation that quantified the status of data rescue, digitalization, and database management across the database categories offers an overview of where systematic biases occur. The significant relationship between data rescue and archiving/management, as well as between digitalization/availability and archiving/management, translates to the current prioritization of uploading specimens when possible while also maintaining databases over time [54, 63]. However, these relationships were heavily influenced by databases already making significant strides towards minimizing taxonomic data gaps, such as citizen science and mixed-collection repositories, while others, such as private and state repositories, may lack the capacity to contribute to the data digitalization effort, despite probable interest and valuable data contributions. These rankings also speak to the importance of database management over time; while collecting and digitizing specimens in the first place is critical to answering questions about biodiversity loss, the maintenance of species records will be critical as taxonomic understanding and species range distributions may shift over time. Despite the tremendous history of biodiversity and taxonomic work nationwide, researchers recognize that a baseline understanding of species diversity is still lacking [60]. The lack of significance between data rescue and digitalization categories highlights the need for continued progress across not only collecting and preserving data, but also providing access to data so that biodiversity targets can be effectively measured. Nevertheless, the continuation of digitized, online platforms such as NHC repositories, in addition to citizen science platforms, can hopefully mitigate some of these taxonomic and data sharing bottlenecks in future research endeavors, and it could even lead to future educational opportunities. Global biodiversity public education programs and efforts to build capacity for communities to engage in broadscale citizen science projects will yield more robust datasets, and with faster rates of data rescue, digitalization, and management due to evolving technological tools [64]. As found through this study, even charismatic taxonomic groups like bumble bees and butterflies will require additional data improvements if collections intend to act as time capsules for future species diversity assessments. Similarly, these improvements will assist in the advancement of sustainability initiatives such as those proposed through the SDGs.

Some limitations of our study included the reliance on records that were considered “complete” (e.g., they included the scientific name and date/location of the collection). Many older, private collections included label names that were illegible to process, and some citizen science records included inaccurate locations that had to be excluded from this effort; these are common issues observed across the respective database types and only emphasize the need for stronger digitalization/availability and archiving/management efforts in the future [3, 14, 21–23]. Additionally, this analysis did not uncover any missing or unexpected species from the databases. While the data availability may be higher for some species taxonomic groups, such as the Nymphalidae or bumble bees, the data available for the parks seems relatively inclusive to what entomologists would expect to observe in the area; or, the species documentation for lesser-known, unexpected, or rare species has fallen through the data rescue cracks, thereby creating unknown data gaps that may never be recovered.

Conclusion

The available data for bumble bee and butterfly species in the GYE varied in their range and were spread across federal, university, and local collections with different digitalization statuses and taxonomic interests. GRTE and YELL are some of the fortunate parks in the US that have a history of insect collections; many others lack data that would make these types of questions difficult to answer [65]. NHC and citizen science databases will continually evolve as more digitalization efforts occur, and this presents ample opportunity for future understanding of species statuses that can benefit learning tools and even conservation targets. Since protected areas are proposed to act as climate refugia for sensitive species diversity given the presumed lower impact of other anthropogenic pressures such as urbanization and habitat degradation [66, 67], this study aimed to determine patterns in bumble bee and butterfly data that have previously been collected, and where there are current taxonomic gaps. As concerns over biodiversity loss amplify and international partners strive to achieve targets set under SDG 15, researchers need to know how tools such as virtual NHC repositories and citizen science data can be informative for species monitoring over time. Targets relevant to biodiversity protection cannot be achieved without a prior understanding of the historic and current data availability and species occurrence status in the complex and varying databases. Therefore, more efforts to prioritize data transparency that minimizes systematic data biases are needed, especially for priority taxa such as pollinators. Together, through resilient partnerships, multiple datasets, and collaborations across protected spaces, we can forge a path towards achieving global sustainability goals and metrics, such as those proposed by SDG 15, securing a flourishing and resilient world of protected habitat for generations to come.

List of abbreviations

UNEP	United Nations Environmental Programme
SDG(s)	Sustainable Development Goal(s)
NHCs	Natural history collections
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
US	United States
ESA	Endangered Species Act
SCAN	Symbiota Collection of Arthropods Network
GBIF	Global Biodiversity Information facility
GRTE	Grand Teton National Park
YELL	Yellowstone National Park
GYE	Greater Yellowstone Ecosystem
NPS	National Park Service
IAR	Investigator annual report

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42055-024-00084-9.

Supplementary Material 1

Acknowledgements
Thank you to Drs. Gillian Bowser, Meena Balgopal, and Emily Fischer from Colorado State University, and Philip Halliwell from Colorado Mountain College for their manuscript advice.

Author contributions

S.W. conceptualized and designed the study. S.W. acquired, analyzed, and intercepted the data. S.W. and S.M. have drafted the work and substantially revised it. All authors read and approved the final manuscript.

Funding

Research is supported by the United States National Science Foundation grants #1624191 and #1645449.

Data availability

The datasets generated and analyzed during the current study are available in the Mendeley Data repository, https://doi.org/10.7632/fgjwsyfbtz.1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 September 2023 / Accepted: 25 March 2024

References

1. Meineke EK, Davies TJ, Daru BH, Davis CC. Biological collections for understanding biodiversity in the Anthropocene. Philos Trans R Soc B. 2019;374(1763):20170386. https://doi.org/10.1098/rstb.2017.0386.
2. Jetz W, McGeoch MA, Guralnick R, Ferrier S, Beck J, Costello MJ, Fernandez M, Geller GN, Keil P, Merow C, Meyer C. Essential biodiversity variables for mapping and monitoring species populations. Nat Ecol Evol. 2019;3(4):539–51. https://doi.org/10.1038/s41559-019-0826-1.
3. Kharouba HM, Lewthwaite JM, Guralnick R, Kerr JT, Vellend M. Using insect natural history collections to study global change impacts: challenges and opportunities. Philos Trans R Soc B. 2019;374(1763):20170405. https://doi.org/10.1098/rstb.2017.0405.
4. Polgar CA, Primack RB, Williams EH, Stichter S, Hitchcock C. Climate effects on the flight period of Lycaenid butterflies in Massachusetts. Biol Conserv. 2013;160:25–31. https://doi.org/10.1016/j.biocon.2012.12.024.
5. Sánchez-Bayo F, Wyckhuys KA. Worldwide decline of the entomofauna: a review of its drivers. Biol Conserv. 2019;232:8–27. https://doi.org/10.1016/j.biocon.2019.01.020.
6. Harvey JA, Tougeron K, Gols R, Heinen R, Abarca M, Abram PK, Basset Y, Berg M, Boggs C, Brodeur J, Cardoso P. Scientists' warning on climate change and insects. Ecol Monogr. 2023;93(1):e1553. https://doi.org/10.1002/ecm.1553.
7. Entomological Society of America. ESA position statement on endangered insect species: protecting endangered insects is in the nation's best interest. Ann Entomol Soc Am. 2018;111(2):81–82.
8. Einhorn C. Are butterflies wildlife? Depends where you live. The New York Times. 2023. https://www.nytimes.com/interactive/2023/03/04/climate/insects-wildlife-us.html. Accessed 18 Mar 2024.
9. Dangles O, Casas J. Ecosystem services provided by insects for achieving sustainable development goals. Ecosyst Serv. 2019;35:109–15. https://doi.org/10.1016/j.ecoser.2018.12.002.
10. United Nations. Goal 15: protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss. Sustainable Development Goals. https://sdgs.un.org/goals/goal15. Accessed 18 Mar 2024.
11. Potts SG, Imperatriz-Fonseca V, Ngo HT, Aizen MA, Biesmeijer JC, Breeze TD, Dicks LV, Garibaldi LA, Hill R, Settele J, Vanbergen AJ. Safeguarding pollinators and their values to human well-being. Nature. 2016;540(7632):220–29. https://doi.org/10.1038/nature20588.
12. Fonseca LM, Domingues JP, Dima AM. Mapping the sustainable development goals relationships. Sustainability. 2020;12(8):3359. https://doi.org/10.3390/su12083359.
13. IPBES. Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination, and food production. In: Potts SG, Imperatriz-Fonseca VL, Ngo HT, Biesmeijer JC, Breeze TD, Dicks LV, Garibaldi LA, Hill R, Settele J, Vanbergen AJ, Aizen MA, Cunningham SA, Eardley C, Freitas BM, Gallai N, Kevan PG, Kovacs-Hostyanszki A, Kwapong PD, Li J, Li X, Martins DJ, Nate-Parra G, Pettis JS, Rader R, Viana BF, editors. Secretariat of the intergovernmental science-policy platform on biodiversity and ecosystem services. Bonn, Germany; 2016. p. 1–36.
14. Cobb NS, Gall LF, Zaspel JM, Dowdy NJ, McCabe LM, Kawahara AY. Assessment of North American arthropod collections: prospects and challenges for addressing biodiversity research. PeerJ. 2019;7(e8086). https://doi.org/10.7717/peerj.8086.
15. Biodiversity Collection Network. Extending U.S. biodiversity collections to promote research and education. Washington, D.C.: American Institute of Biological Sciences; 2019. p. 1–8.
16. Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience. 2009;59(11):977–84. https://doi.org/10.1525/bio.2009.59.11.9.
17. Forister ML, McCall AC, Sanders NJ, Fordyce JA, Thorne JH, O'Brien J, Waetjen DP, Shapiro AM. Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. Proc Natl Acad Sci. 2010;107(5):2088–92. https://doi.org/10.1073/pnas.0909686107.
18. Mazumdar S, Ceccaroni L, Piera J, Hölker F, Berre A, Arlinghaus R, Bowser A. Citizen science technologies and new opportunities for participation. UCL Press. https://doi.org/10.14324/111.9781787352339.
19. National Academies of Sciences, Engineering, and Medicine (NAS). Learning through citizen science: enhancing opportunities by design. Washington, DC: The National Academies Press; 2018. https://doi.org/10.17226/25183.
20. Ellwood ER, Crimmins TM, Miller-Rushing AJ. Citizen science and conservation: recommendations for a rapidly moving field. Biol Conserv. 2017;208:1–4. https://doi.org/10.1016/j.biocon.2016.10.014.
21. Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK. Next steps for citizen science. Science. 2014;343(6178):1436–37. https://doi.org/10.1126/science.1251554.
22. Lukyanenko R, Parsons J, Wiersma YF. Emerging problems of data quality in citizen science. Conserv Biol. 2016;30(3):447–49. https://doi.org/10.1111/cobi.12706.
23. Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE. The accuracy of citizen science data: a quantitative review. Bull Ecol Soc Am. 2017;98(4):278–90. https://doi.org/10.1002/bes2.1336.
24. Mondejar ME, Avtar R, Diaz HL, Dubey RK, Esteban J, Gómez-Morales A, Hallam B, Mbungu NT, Okolo CC, Prasad KA, She Q. Digitalization to achieve sustainable development goals: steps towards a Smart Green planet. Sci Total Environ. 2021;794:148539. https://doi.org/10.1016/j.scitotenv.2021.148539.
25. Balogun AL, Marks D, Sharma R, Shekhar H, Balmes C, Maheng D, Arshad A, Salehi P. Assessing the potentials of digitalization as a tool for climate change adaptation and sustainable development in urban centres. Sustainable Cities Soc. 2020;53:101888. https://doi.org/10.1016/j.scs.2019.101888.
26. Rocha-Ortega M, Rodriguez P, Córdoba-Aguilar A. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. Ecol Entomol. 2021;46(4):718–28. https://doi.org/10.1111/een.13027.
27. Koffler S, Barbiéri C, Ghilardi-Lopes NP, Leocadio JN, Albertini B, Francoy TM, Saraiva AM. A buzz for sustainability and conservation: the growing potential of citizen science studies on bees. Sustainability. 2021;13(2):959. https://doi.org/10.3390/su13020959.
28. Brooks SJ, Self A, Toloni F, Sparks T. Natural history museum collections provide information on phenological change in British butterflies since the late-nineteenth century. Int J Biometeorol. 2014;58:1749–58. https://doi.org/10.1007/s00484-013-0780-6.
29. Ansari NA, Agus C, Nunoo EK. Foundations of 'SDG15–LIFE on land': earth, forests and biodiversity. In: SDG15–life on land: towards effective

biodiversity management. Emerald Publishing Limited; 2021. p. 7–48. https://doi.org/10.1108/978-1-80117-814-320211004.
30. Hardisty AR, Ellwood ER, Nelson G, Zimkus B, Buschbom J, Addink W, Rabeler RK, Bates J, Bentley A, Fortes JA, Hansen S. Digital extended specimens: enabling an extensible network of biodiversity data records as integrated digital objects on the internet. BioScience. 2022;72(10):978–87. https://doi.org/10.1093/biosci/biac060.
31. Holling CS. Resilience and stability of ecological systems. Annu Rev Ecol Syst. 1973;4(1):1–23. https://doi.org/10.1146/annurev.es.04.110173.000245.
32. Gunderson LH, Holling CS, editors. Panarchy: understanding transformations in human and natural systems. Island press; 2002.
33. Winfree R. Global change, biodiversity, and ecosystem services: what can we learn from studies of pollination? Basic Appl Ecol. 2013;14(6):453–60. https://doi.org/10.1016/j.baae.2013.07.004.
34. Bodin P, Wiman B.Resilience and other stability concepts in ecology: notes on their origin, validity, and usefulness. ESS Bulletin. 2004;2(2):33–43.
35. Gunderson LH. Ecological resilience—in theory and application. Annu Rev Ecol Syst. 2000;31(1):425–39. https://doi.org/10.1146/annurev.ecolsys.31.1.425.
36. GBIF.org. GBIF occurrence download for Nymphalidae. 2022a. https://doi.org/10.15468/dl.6ptczh.
37. GBIF.org. GBIF occurrence download for Lycaenidae. 2022b. https://doi.org/10.15468/dl.mms34y.
38. GBIF.org. GBIF occurrence download for Hesperiidae. 2022c. https://doi.org/10.15468/dl.7sdefs.
39. GBIF.org. GBIF occurrence download for Pieridae. 2022d. https://doi.org/10.15468/dl.fpjk8z.
40. GBIF.org. GBIF occurrence download for Papilionidae. 2022e. https://doi.org/10.15468/dl.4wtm9r.
41. GBIF.org. GBIF occurrence download for *Bombus*. 2022f. https://doi.org/10.15468/dl.jdxpnh.
42. Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. Darwin core: an evolving community-developed biodiversity data standard. PLoS One. 2012;7(1):e29715. https://doi.org/10.1371/journal.pone.0029715.
43. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686. https://doi.org/10.21105/joss.01686.
44. Whipple S. GYE digitalization data. Mendeley Data. 2023;V1. https://doi.org/10.17632/fgjwsyfbtz.1.
45. Hill MO. Diversity and evenness: a unifying notation and its consequences. Ecology. 1973;54(2):427–32. https://doi.org/10.2307/1934352.
46. Smith B, Wilson JB. A consumer's guide to evenness indices. Oikos. 1996;70–82. https://doi.org/10.2307/3545749.
47. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Oksanen MJ.Package 'vegan'. Commun Ecol Package Vers. 2013;2(9):1–295.
48. Strong WL. Biased richness and evenness relationships within Shannon–Wiener index values. Ecol Indic. 2016;67:703–13. https://doi.org/10.1016/j.ecolind.2016.03.043.
49. Freitas AV, Brown Jr KS. Phylogeny of the nymphalidae (Lepidoptera). System Biol. 2004;53(3):363–83. https://doi.org/10.1080/10635150490445670.
50. Munroe E. The classification of the Papilionidae (Lepidoptera). Mem Ent Soc Can. 1960;92(S17):5–1. https://doi.org/10.4039/entm9217fv.
51. Auckland JN, Debinski DM, Clark WR. Survival, movement, and resource use of the butterfly *Parnassius clodius*. Ecol Entomol. 2004;29(2):139–49. https://doi.org/10.1111/j.0307-6946.2004.00581.x.
52. Caruthers JC, Debinski DM. Montane meadow butterfly species distributions in the Greater Yellowstone Ecosystem. Yellowstone Ecosyst Rep. 2006;30(14).
53. Hines HM. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: apidae: bombus). System Biol. 2008;57(1):58–75. https://doi.org/10.1080/10635150801898912.
54. Shirey V, Belitz MW, Barve V, Guralnick R. A complete inventory of North American butterfly occurrence data: narrowing data gaps, but increasing bias. Ecography. 2021;44(4):537–47. https://doi.org/10.1111/ecog.05396.
55. Kim KC, Byrne LB. Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. Ecol Res. 2006;21:794–810. https://doi.org/10.1007/s11284-006-0035-7.
56. Bertrand Y, Pleijel F, Rouse GW. Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. Syst Biodivers. 2006;4(2):149–59. https://doi.org/10.1017/S1477200005002908.
57. Bachmann N, Tripathi S, Brunner M, Jodlbauer H.The contribution of data-driven technologies in achieving the sustainable development goals. Sustainability. 2022;14(5):2497.
58. Corbane C, Pesaresi M, Politis P, Syrris V, Florczyk AJ, Soille P, Maffenini L, Burger A, Vasilev V, Rodriguez D, Sabo F.Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. Big Earth Data. 2017;1(1-2):118–44.
59. Nilashi M, Keng Boon O, Tan G, Lin B, Abumalloh R. Critical data challenges in measuring the performance of sustainable development goals: solutions and the role of big-data analytics. Harvard Data Sci Rev. 2023;5(3).
60. Kass JM, Guénard B, Dudley KL, Jenkins CN, Azuma F, Fisher BL, Parr CL, Gibb H, Longino JT, Ward PS, Chao A. The global distribution of known and undiscovered ant biodiversity. Sci Adv. 2022;8(31):eabp9908. https://doi.org/10.1126/sciadv.abp9908.
61. Anuardo RG, Espuny M, Costa AC, Espuny AL, Kazançoğlu Y, Kandsamy J, de Oliveira OJ. Transforming E-waste into opportunities: driving organizational actions to achieve sustainable development goals. Sustainability. 2023;15(19):14150. https://doi.org/10.3390/su151914150.
62. Pérez-Martínez J, Hernandez-Gil F, San Miguel G, Ruiz D, Arredondo MT. Analysing associations between digitalization and the accomplishment of the sustainable development goals. Sci Total Environ. 2023;857:159700.
63. Shirey V. Visualizing natural history collection data provides insight into collection development and bias. Biodivers Data J. 2018;6. https://doi.org/10.3897/BDJ.6.e26741.
64. Callaghan CT, Rowley JJ, Cornwell WK, Poore AG, Major RE.Improving big citizen science data: moving beyond haphazard sampling. PLoS Biol. 2019;17(6):e3000357.
65. Shafer CL. National park and reserve planning to protect biological diversity: some basic elements. Landscape Urban Plann. 1999;44(2-3):123–53. https://doi.org/10.1016/S0169-2046(98)00115-7.
66. Morelli TL, Daly C, Dobrowski SZ, Dulen DM, Ebersole JL, Jackson ST, Lundquist JD, Millar CI, Maher SP, Monahan WB, Nydick KR. Managing climate change refugia for climate adaptation. PLoS One. 2016;11(8):e0159909. https://doi.org/10.1371/journal.pone.0519909.
67. Barrows CW, Ramirez AR, Sweet LC, Morelli TL, Millar CI, Frakes N, Rodgers J, Mahalovich MF. Validating climate-change refugia: empirical bottom-up approaches to support management actions. Front Ecol Environ. 2020;18(5):298–306. https://doi.org/10.1002/fee.2205.

Publisher's Note