Journal *of* Melittology
Bee Biology, Ecology, Evolution, & Systematics
No. 123(3), 17–77 8 December 2025

Improving the standardization of wild bee occurrence data: Towards a formal wild bee data standard

Brianne Du Clos[1], Katja C. Seltmann[2], Nash E. Turley[3], Clare Maffei[4], Erika M. Tucker[5], Ian G. Lane[6], Hannah K. Levenson[7], & S. Hollis Woodard[8]

Abstract. Conservation and management of wild bees is hindered by the variety of ways wild bee occurrence data are recorded, managed, and shared. Here, we present solutions to address this issue and introduce *The Wild Bee Data Standard*, a standardized means of recording and reporting data associated with wild bee occurrences, including physical specimens and photo observations. This standard aligns with contemporary data management practices widely adopted by the broader biodiversity data community. We propose a set of terms for the standard that describe various features of bee occurrences, including collection method and location, taxonomic verification, and final record storage. We emphasize the importance of providing sampling protocol and effort information with wild bee occurrence data and offer guidance to make this a more common practice. We describe how to translate data not currently aligned with the standard to meet its conditions, and how to upload those data to an accessible online repository. We provide case studies, data entry templates, a glossary of terms, and additional resources to guide new users to implementing the standard. We also present a forum, established as a GitHub repository, to support continued development of the standard. Recognizing the significant change this represents for current data practices, we outline the benefits for the bee research and conservation community that will result from improved data standards. We advocate for making all historical, current, and future bee occurrence data openly available to facilitate more rigorous and comprehensive research, conservation, and management of wild bees. This contribution is part of a series developed in association with the U.S. National Native Bee Monitoring Network to standardize bee monitoring practices.

[1] Department of Entomology, University of California, Riverside, Riverside, CA, 92521, USA Current address: Louisiana Universities Marine Consortium, Chauvin, LA, 70344, USA (bduclos@lumcon.edu)
[2] Cheadle Center for Biodiversity and Ecological Restoration, University of California, Santa Barbara CA, 93106, USA (seltmann@ucsb.edu)
[3] Department of Entomology, The Pennsylvania State University, University Park, PA, 16802, USA (nqt5263@psu.edu)
[4] U.S. Fish and Wildlife Service, Laurel, MD, 20708, USA (clare_maffei@fws.gov)
[5] Biodiversity Outreach Network, Flagstaff, AZ, 86001, USA (emtuckerlab@gmail.com)
[6] U.S. Fish and Wildlife Service, National Wildlife Refuge Program, Bloomington, MN, 55425, USA (ian_lane@fws.gov)
[7] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, 27695, USA (hklevens@ncsu.edu)
[8] Department of Entomology, University of California, Riverside, Riverside, CA, 92521, USA (hollis.woodard@ucr.edu)

ISSN 2325-4467

INTRODUCTION

Wild bee occurrence data are generated when bees are collected or observed in the field, and include the date collected or observed, collector or observer identity, geographic location, and other information. Occurrence data generated from specimens housed in collections are foundational to biological research (Meineke *et al*., 2018; Nachman *et al*., 2023). Occurrence data are increasingly being generated by field observations of live organisms, which can supplement collections-based data (Briggs *et al*., 2022; Boone *et al*., 2023; MacPhail *et al*., 2024) but are currently suitable only for a subset of bee species (Turley *et al*., 2024). Occurrence data are most useful when digitized (Cobb *et al*., 2019), annotated (Rousseau *et al*., 2024), georeferenced (Seltmann *et al*., 2018), and made openly available (Nelson & Ellis 2018), yet these conditions are often unmet, which limits reproducibility (Turney *et al*., 2015; Packer *et al*., 2018). Openly available occurrence data can be viewed and downloaded online from a number of biodiversity data portals, including the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), Integrated Digitized Biocollections (iDigBio; https://www.idigbio.org/portal/search), iNaturalist (https://www.inaturalist.org/), the Symbiota Collection of Arthropods Network (SCAN; https://scan-bugs.org/portal/), Ecdysis (https://ecdysis.org), and the Bee Library (https://library.big-bee.net). These data portals typically provide data following FAIR principles (Findable, Accessible, Interoperable, and Reusable; Wilkinson *et al*., 2016).

Adhering to FAIR principles can be accomplished in part by following data standards, or agreed-upon terms that describe data points or datasets. The Darwin Core data standard (DwC; https://dwc.tdwg.org/) includes a set of standardized terms that can be applied to any biodiversity dataset as column headings in spreadsheets or delimited files. The terms are linked to definitions about the data, providing a consistent way of sharing biodiversity data (Wieczorek *et al*., 2012). Darwin Core is managed by Biodiversity Information Standards (TDWG), previously known as the Taxonomic Databases Working Group, and is used by leading biodiversity data providers, including GBIF, iDigBio, and iNaturalist. The Biotic Observation Minimum Specification (BOMS) is a data standard in development at the U.S. Fish and Wildlife Service for use in inventory and monitoring work on National Wildlife Refuges that is aligned with Darwin Core (U.S. Fish and Wildlife Service, 2023). Here, we used Darwin Core terms, and drew influence from BOMS, to create *The Wild Bee Data Standard* v.1.0.0 (Appendix 1). When working with DwC terms, whether by creating new datasets or translating existing data, not all terms need to be used; therefore, *The Wild Bee Data Standard* only includes DwC terms that are most relevant to wild bee occurrence-based analyses. Many of these terms describe information that is familiar to those working with wild bee occurrence data, including date, time, location, and species collected, but we also emphasize reporting information on sampling methods and study design (Montgomery *et al*., 2021; Rousseau *et al*., 2024; Levenson *et al*., 2025b). Some DwC terms recommend using controlled vocabularies, which are lists of allowable entries that constrain data entries for particular types of information. Controlled vocabulary lists for DwC terms are provided as needed in *The Wild Bee Data Standard*.

Also important for FAIR data is the creation and inclusion of detailed metadata, or information that describes the data, which can also be provided following a set of standards (Bloom *et al*., 2021). Ecological Metadata Language (EML) is typically used in biodiversity data management, with standard methods for reporting the source information for occurrence datasets, including the project name, data creator(s), and

data provider(s), along with the type of data being provided, how it should be cited, and other descriptive information (Jones *et al.*, 2019).

In this contribution, we aim to clarify how to create, find, access, format, share, and use openly available wild bee occurrence data and metadata to promote FAIR principles and advance wild bee species conservation and management. We propose implementing *The Wild Bee Data Standard* and provide a framework for doing so. We provide a glossary of terms (Appendix 2) and a list of additional resources (Appendix 3) to supplement this manuscript. This standard was developed as part of the U.S. National Native Bee Monitoring Research Coordination Network, hereafter referred to as the Bee Monitoring RCN. Although *The Wild Bee Data Standard* was developed to support standardized protocols for wild bee monitoring written by the Bee Monitoring RCN (Cariveau *et al.*, 2025; Levenson *et al.*, 2025a; López-Uribe *et al.*, 2025; Otto *et al.*, 2025; Strange *et al.*, 2025), any project that collects information on wild bee occurrence, particularly with physical specimens or photo observations, can use the standard to plan data collection, digitize data, and ultimately, make data openly available. Those who generate wild bee occurrence data often gather similar variables (*i.e.*, species name, sex, date, location, collector), but may be naming and managing those variables differently. Adhering to the *The Wild Bee Data Standard* and the data management practices outlined here ensures that wild bee occurrence data collected for any project can be synthesized and used by various entities for more thorough and rigorous analyses. For example, analyses relying on the knowledge of wild bee species' ranges and population statuses, which are a cornerstone of conservation, are currently severely limited by the availability and suitability of openly available wild bee data (Graves *et al.*, 2020; Chesshire *et al.*, 2023; Rousseau *et al.*, 2024; 2025). We thus advocate for making all historical, current, and future wild bee occurrence data openly available following FAIR principles and *The Wild Bee Data Standard*, which will ultimately lead to more effective management and conservation actions for wild bees. Although this goal is clear, we anticipate the following obstacles to this request and provide solutions for each:

First, some historical data may not have associated information on sampling effort and protocol to report. We emphasize these two pieces of information in *The Wild Bee Data Standard*, as they allow more rigorous conservation status assessments and other quantitative analyses of wild bee occurrence (Bloom *et al.*, 2021; Rousseau *et al.*, 2024). Even if sampling effort and protocol information are not available, occurrence data should still be made openly available, as it is valuable for other purposes, such as determining potential ranges of wild bee species and population shifts over time (Burkle *et al.*, 2013; Mathiasson & Rehan, 2019; Graham *et al.*, 2021). Future collecting efforts should improve documentation on sampling effort and protocols.

Second, most digitized datasets will not follow *The Wild Bee Data Standard* in their current form. Conversion to Darwin Core terms (a process known as mapping) to conform to *The Wild Bee Data Standard* can be done by using one of the accompanying templates (Du Clos *et al.*, 2025) or a GBIF template, by updating existing in-house templates, or as part of the publication process to a platform that exports datasets in the Darwin Core Archive format that is usable by GBIF. There are multiple platforms that have built in mapping tools to help with this conversion, including the GBIF Integrated Publishing Toolkit (IPT; Robertson *et al.*, 2014) and Collections Management Systems (CMS, which we elaborate on below). Additionally, photo observation data can be uploaded to iNaturalist, which provides research-grade occurrence data to GBIF following Darwin Core standards, effectively mapping the data for uploaders.

Third, we recognize that some data cannot be made openly available or must be masked due to privacy, security, conservation-related, or other concerns. We advise that data providers use the Darwin Core term **dwc:informationWithheld** to indicate that the precise location, full taxonomic name, or other information has not been shared owing to these concerns. This allows for transparency of occurrence information, alignment with *The Wild Bee Data Standard*, and a path towards contacting data providers for special uses of data, while protecting threatened or endangered species. We recommend reviewing current data masking policies and engaging partners in discussions about data transparency in an effort to bring more data online.

METHODS FOR SHARING AND STANDARDIZING WILD BEE OCCURRENCE DATA

STANDARDIZING CURATION OF WILD BEE DATA AND PHYSICAL SPECIMENS. Curating wild bee occurrence data involves the preparation, handling, and storage of digital records and physical specimens. Following *The Wild Bee Data Standard*, we advise digitizing field-collected data with selected Darwin Core terms as column headings. This promotes interoperability, or the ability to exchange information across platforms, and improved data understanding between users. We provide templates aligned with *The Wild Bee Data Standard* to simplify digitization and mapping of currently existing datasets (Du Clos *et al*., 2025).

Occurrence data can be extended beyond a specimen or an observation to include additional data about the occurrence (*e.g.*, image data of the specimen or its habitat, genetic data, interaction or floral use data; Lendemer *et al*., 2020). Sharing occurrence data and any related information is crucial, as these extended resources are critical for confirming specimen identifications, understanding bee functional traits, and enhancing our comprehension of environmental influences on bee abundance and diversity. Tools for the generation, management, and sharing of image and functional trait data are being developed by the Big-Bee project (https://big-bee.net; see also Ostwald *et al*., 2024). We provide standards for both image data and floral use data in *The Wild Bee Data Standard*. We also provide standards for bee occurrence data to be used in molecular (López-Uribe *et al*., 2025) and pathogen analyses (Strange *et al*., 2025).

Improving the curation of wild bee physical specimens for identification, long-term storage, and extraction of new and derived data (*e.g.*, molecular work, physical measurement, pollen collection) is essential for wild bee occurrence data to more accurately inform our understanding of bee distribution, population trends, conservation status, and other relevant metrics. Specimens, particularly those collected in a bowl or cup trap, should be cleaned following guidance from *The Very Handy Bee Manual* (A Collective 2024). Specimens caught with a net may not need to be cleaned, as they may have pollen or pathogens that could be collected for identification or further study. Specimens should be pinned neatly so that all external anatomy can be seen, and internal anatomy can be accessed if necessary and labeled with the date, coordinates, and name(s) of the collector(s) (Burrows *et al*., 2021; A Collective 2024).

Once prepared for curation, and either before or after identifications have been verified, specimens are best placed in a public natural history collection (*i.e.*, accessioned), allowing access to the specimens for taxonomic study and research purposes in perpetuity. Collections would ideally be located within museums or other established collection facilities; however, it is often difficult to accomplish museum

accession owing to the limited space and infrastructure to support current collections (Cobb *et al*., 2019). To improve the management and curation of wild bee physical specimen collections in the US, we need more collection facilities to store specimens and more collection management personnel to support these specimens. Collection management personnel have a vast suite of responsibilities including, but not limited to, proper specimen care and storage, digitizing specimens for in-house and openly accessible databases, managing specimen databases, managing specimen status (*i.e*., accessions, acquisitions, loans, de-accessions), coordinating with visiting scientists for access to those specimens, and conducting public outreach related to the collections. Natural history collections of wild bee specimens are foundational to pollination biology, ecology, and conservation; their importance cannot be overstated and increased and sustained support is critical for their continuity (Bartomeus *et al*., 2019).

Standardizing data access. Making occurrence data openly available has multiple benefits to wild bee research, conservation, and policy (Rousseau *et al*., 2024). To be openly available, data must be downloadable by users without explicit permission from the provider (*i.e*., users do not need to email the provider to obtain a dataset), and the conditions for using the data are made clear by the provider. Conditions for reuse are typically determined by adding a use license, such as a Creative Commons license, to the data when published online. In order for data to be openly available, they first need to be digitized, or converted into a digital format using an electronic device with internet connection capabilities (*e.g*., computer, tablet, phone, etc.) (Nelson & Ellis, 2018). Given the rich history of insect specimen collection, there are an estimated 4.7 million undigitized bee specimens across the US that could contribute valuable information on long-term wild bee population trends (Chesshire *et al*., 2023). Digitizing these would more than double the number of bee records available for the US, as 2.9 million digitized bee records were acquired after an extensive search across multiple platforms by Chesshire *et al*. (2023). A primary funding source for collections digitization efforts, Advancing the Digitization of Biological Collections (ADBC), was recently archived by the National Science Foundation (NSF), and collections funding was moved to a more general infrastructure program, called the Capacity Program. One of the last ADBC-funded projects focused on digitizing wild bee specimens was Big-Bee (DBI#2102006; https://big-bee.net), whereas another wild bee digitization project, iDigBees (DBI#2216927; https://idigbees.org), was funded through the Capacity Program. We strongly support these projects and advocate for continued funding support for bee specimen digitization.

Digitizing field-collected data into one of the templates accompanying *The Wild Bee Data Standard* (Du Clos *et al*., 2025) facilitates uploading to a data aggregator, CMS, or other web-based data platforms that adhere to the Darwin Core data standard. One barrier to making occurrence data openly available is selecting a platform for data sharing. Here, we highlight three platforms (Fig. 1) and provide guidance for uploading data to each, while recognizing that other valid options could also be pursued. We are not advocating for all data providers to follow the same data sharing pathway or to use the same platform; if you are a data provider with established data sharing practices following Darwin Core standards, we encourage you to continue those practices. Our guidance here is intended for those who do not have these practices in place and are looking for a starting point to begin sharing their data. Our goal is for all historical, current, and future wild bee occurrence data to be openly available following FAIR

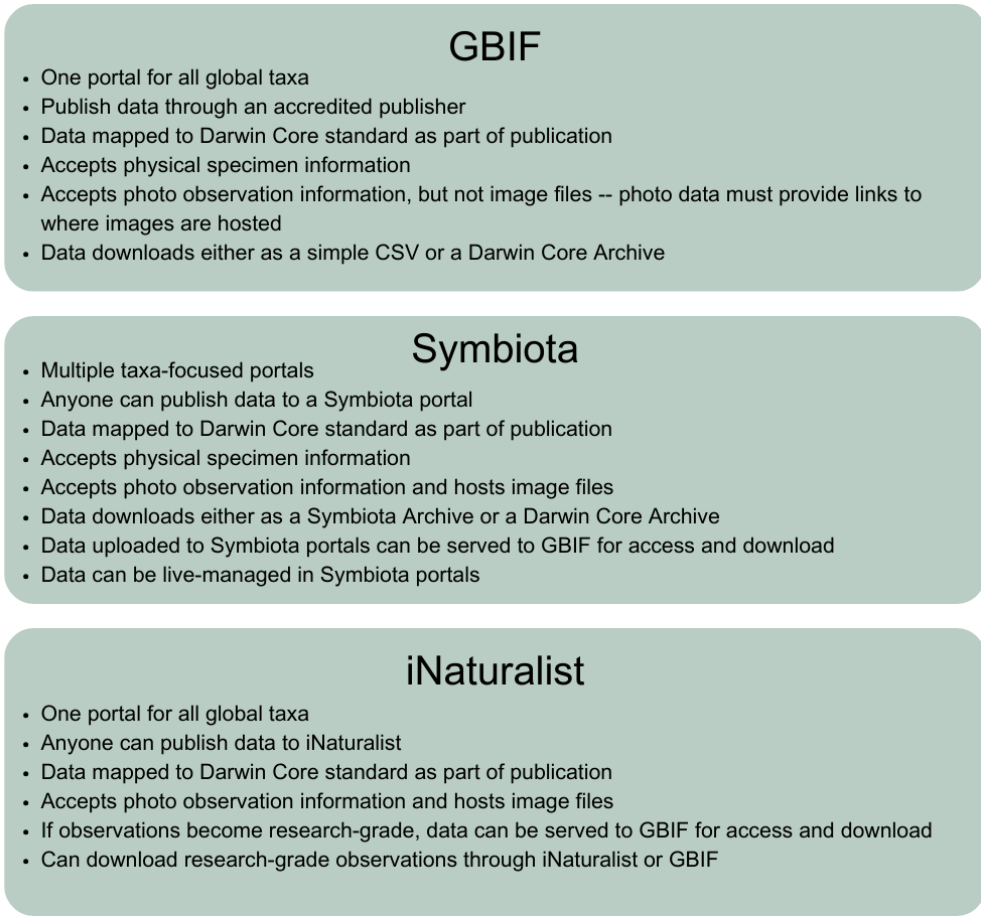

Figure 1. Details of three platforms that provide openly available wild bee occurrence data following FAIR data principles. Occurrence data can be downloaded from all three platforms; however, Symbiota and iNaturalist also send data to GBIF, meaning that data from all three platforms can be accessed in one place. This also means that providers unable to publish data directly to GBIF can use either Symbiota or iNaturalist to get data there.

principles and *The Wild Bee Data Standard*, which will ultimately lead to more effective management and conservation actions for wild bees.

The first option is GBIF, a data aggregator that publishes physical specimen and photo observation occurrence data through its IPT and makes them openly available for download. The GBIF IPT implements FAIR principles by mapping occurrence data to Darwin Core terms as part of the publishing process and producing an accompanying metadata document following Ecological Metadata Language (EML, Robertson *et al*., 2014; Jones *et al*., 2019). It is worth noting that familiarity with GBIF can vary by employment sector, with government employees often using other platforms to share and download occurrence data (Martín-Mora *et al*., 2020). Additionally, data can only be published to GBIF through accredited publishers. This can present a barrier to data providers looking to make their occurrence data openly available, if they are not already connected with a publisher. However, with dedicated digitization efforts and effective communication involving new or existing accredited publishers, the

awareness and use of GBIF for sharing data could grow, allowing data from multiple agencies and sectors to be published online.

The second option for data publishing is Symbiota (https://symbiota.org; Gries *et al.*, 2014), an open-source CMS supporting taxon-focused data portals that manage and share occurrence data, including physical specimens, photo observations, and published datasets. As Symbiota portals host multiple datasets and make them openly accessible for download, these portals also function as data aggregators. Any data provider can publish occurrence data directly to a Symbiota portal. Within the portals, data providers have the option to edit and publish updated datasets through a web-accessible databasing infrastructure; these datasets are called "live data" See https://ecdysis.org/collections/misc/collprofiles.php?collid=120 for an example of live wild bee occurrence data on a Symbiota portal. Data providers can also choose to publish data managed elsewhere to a Symbiota portal; these datasets are called "snapshots". See https://library.big-bee.net/portal/collections/misc/collprofiles.php?collid=34 for an example of a snapshot dataset on a Symbiota portal. Symbiota portals that manage and share insect data include Bee Library, SCAN, and Ecdysis. These portals support digitization and data sharing efforts, in part by providing data management and quality control tools, but also through a robust community of users, providers, and developers. Importantly, Symbiota portals can publish data directly to GBIF (https://docs.symbiota.org/Collection_Manager_Guide/Data_Publishing/publishing_gbif/; see https://www.gbif.org/dataset/a42d07b3-e34d-4a65-b7e3-6aafa9f8f27b for an example of live data from a Symbiota portal published to GBIF). Other CMS that manage occurrence data include Arctos (https://arctosdb.org/; Cicero *et al.*, 2024) and Specify (https://www.specifysoftware.org/). Like Symbiota, these platforms support the management of physical specimen and photo observation data through web-accessible databasing infrastructure, have engaged user communities and accessible developers, and can publish or format these data for publication to GBIF. Arctos is also a data aggregator (https://arctos.database.museum/), but Specify does not currently aggregate data. Thus, if you are currently using other CMS to manage and share data, then we advise continuing to do so.

The third option is iNaturalist. For current and future projects involving photo observations only, iNaturalist can be used to share occurrence data following FAIR principles if verifiable observations that provide a date, latitude and longitude coordinates, and a photo of an organism that is not captive or cultivated become research-grade. An observation becomes research-grade if more than ⅔ of identifiers agree on a species-level taxonomic identification (Campbell *et al.*, 2023). Providing a suggested identification when uploading observations can increase the visibility of records, making it more likely that community identifiers provide the verifications to achieve research-grade status. Data can be downloaded directly from iNaturalist, and a subset (depending on the license applied to the observation) of research-grade observations are pushed to GBIF following Darwin Core standards. We advise only using research-grade records if using iNaturalist data for statistical analyses.

Promoting data accuracy. Data shared in online repositories can have taxonomic or geographic inaccuracies (Orr *et al.* 2021; Chesshire *et al.*, 2023; Dorey *et al.*, 2023). When conducting analyses with wild bee occurrence data downloaded from online biodiversity data platforms, cleaning the data (*i.e.*, removing erroneous data and updating data fields as appropriate) to improve their accuracy is of critical importance.

The statistical software R (R Core Team 2024) has an add-on package called BeeBDC (Dorey *et al*., 2023) that cleans occurrence data to provide more accurate taxonomic and geographic information. We strongly advise either cleaning downloaded data using BeeBDC before any analysis or using the cleaned wild bee occurrence dataset provided in Dorey *et al*. (2023) to conduct analyses relying on bee species locations and validated identifications. BeeBDC can also be used to clean data before uploading, as can GBIF and some other CMS in a more limited fashion, further reducing the propagation of erroneous data for future users. Following these methods helps prevent spurious conclusions that can arise from inaccurate data.

CREDITING DATA SOURCES. It is important to acknowledge and credit everyone who contributed to a shared dataset, whether they collected, prepared, or managed the occurrence data, or collaborated on the project for which the occurrence data were collected (Duke & Porter, 2013). There are multiple ways to accomplish this. Here, we outline practices that are simple to adopt and standardize across shared datasets, in order of preference. When using shared data, give appropriate credit to those involved in generating and managing the data by:

Following proper citation procedures for data downloaded from portals. GBIF provides clear citation guidelines (https://www.gbif.org/citation-guidelines), the most important of which is to *cite the Digital Object Identifier (DOI) of any data downloads in work involving those data*. We cannot emphasize the importance of citing DOIs enough; it is a critical practice that must be adopted by all users of occurrence datasets downloaded online. GBIF creates DOIs for every data download; sharing the DOI allows end users to access the same dataset(s) used in published works. DOIs can be provided for individual datasets or groups of datasets, which are called derived datasets by GBIF. For example, https://doi.org/10.15468/6autvb is the DOI for Droege & Maffei (2025), and https://doi.org/10.15468/dl.6cxfsw is the DOI for most of the data used in Chesshire *et al.* (2023); see Chesshire (2023). Always cite DOIs for derived datasets generated from search filters; for individual datasets, see below. Symbiota provides general citation guidelines (https://docs.symbiota.org/Collection_Manager_Guide/data_citations); each Symbiota portal can customize these guidelines to suit their needs. At a minimum, the portal itself should be cited. Further, each dataset downloaded can be listed in the citation. This is feasible for a small number of citations; for large lists of datasets, use a supplementary file (see below).

Including names of personnel and institutions in data fields. It is important to avoid ambiguity when naming personnel and institutions in occurrence data, especially if there are similar names. To credit bee collectors, use the term **dwc:recordedBy** and list their full name(s) (when known) for each occurrence in the dataset. For additional detail, if possible, list an ORCID iD (https://orcid.org/; Haak *et al*., 2012) for the collector under the term **dwc:recordedByID**. To credit bee identifiers, use the term **dwc:identifiedBy** and list their full name(s) (when known) for each occurrence in the dataset. For additional details, if possible, list an ORCID iD for the identifier under the term **dwc:identifiedByID**. This allows users of the published data to clearly determine the personnel involved in generating occurrence records; further, this information can then be integrated into the online database Bionomia (see below). The term **dwc:georeferencedBy** can be used to cite personnel who handled spatial location of the data. Indicating where the data are housed using the terms **dwc:institutionCode** and/or **dwc:collectionCode** acknowledges the role of the hosting institutions and promotes

access to specimens for further study. If possible, provide a unique identifier using the terms **dwc:institutionID** and **dwc:collectionID**; these identifiers may be found in the Global Registry of Scientific Collections (GrSciColl; https://scientific-collections.gbif.org).

Promoting the use of Bionomia (https://bionomia.net; Shorthouse, 2020) to aggregate digitized identified specimens to identifier(s) and collector(s). Bionomia is an online database that connects specimen occurrence records from GBIF to identifier(s) and collector(s) using unique identifiers, allowing for those personnel to develop an online profile of specimens they identified or collected. Bionomia downloads these records from GBIF every two weeks, and any records with content in the **dwc:identifiedBy,dwc:identifiedByID, dwc:recordedBy,** or **dwc:recordedByID** columns are reviewed, resolved, and published to the website. Identifiers or collectors can claim Bionomia profiles with their ORCID iDs (*e.g.*, https://bionomia.net/0000-0002-8822-2315) and create a DOI for the dataset displayed in their profile that becomes citable and trackable; see Seltmann (2022) for an example of a dataset generated by Bionomia. Note, however, that Bionomia does not collect information on photo observations.

Citing datasets and acknowledging data curators by name in published work. When using data from a platform such as GBIF or a Symbiota portal, a list of all datasets used must be provided. When downloading from GBIF, this list is provided for you (see GBIF, 2021). If there are datasets that comprise more than 5% of a download, providers of those datasets must be specifically named and cited in published work. An example of this can be found in the Acknowledgements section of Rousseau *et al*. (2024): "*A big thank you to all data owners and providers, and especially to the following, who provided at least 5% of the records used in the dataset: American Museum of Natural History, iNaturalist, University of Kansas Biodiversity Institute, U.S. Department of Agriculture, and U.S. Geological Survey; we also cite the providers associated with these collections in our references*." Alternatively, personnel associated with the datasets can be named. This idea is adopted from the USGS Bird Banding Laboratory Data Release Policy and can be adopted through wild bee research as well. In addition to naming and citing large dataset providers, citations for all downloaded datasets must be listed in either a table in the main body text (see Levenson *et al*., 2024) or a supplementary file. If using a supplementary file to list these dataset citations, that file must then be referenced in the main body text of any published work.

The Contributor Role Taxonomy (CRediT) includes a role for Data Curation, defined as "management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use" (Brand *et al*., 2015). Co-authors who curate data used in analyses can be acknowledged in a CRediT statement included in any published work. We recognize, however, that curatorial activities are vast and CRediT may be too broad for sufficient acknowledgement. We advocate for continued development of standardized acknowledgement methods for those who contribute to and curate shared occurrence datasets (Thessen *et al*., 2019). More broadly, we advise crediting or acknowledging the stewards, managers, or owners of the land where wild bees are observed or collected for research purposes. This can be done in the Acknowledgements section of publications relying on such data.

Formalizing data sharing policies and preparing for future data policy shifts. Beyond concern for proper credit, scientists may hesitate to share data because they are concerned about how their data may be used. We advise adopting language from one

or multiple existing documents on data ethics, data integrity, data sovereignty, data agreements, and other guidance to create a Code of Conduct or Norms for Data Use for using data following *The Wild Bee Data Standard*. Examples include the USGS Bird Banding Laboratory Data Release Policy, the CARE Principles for Indigenous Data Governance (Carroll *et al*., 2020), the US Federal Data Strategy Data Ethics Framework, the Canadensys Norms for Data Use, and the VertNet Norms for Data Use. Links to all of these examples are provided in Appendix 3. These documents typically describe a series of behaviors that data users should understand before downloading and using any data, including proper use cases, crediting practices, and caveats related to data accuracy.

Further, to ensure data used in publications can be located online, we advise the use of data availability statements in published manuscripts, which are now required to accompany peer-reviewed publications in many scientific journals. A data availability statement is a short, straightforward statement that describes where the data supporting a peer-reviewed publication can be found. An example that would represent data shared following *The Wild Bee Data Standard* is: "Data and metadata are aligned with *The Wild Bee Data Standard* (citation) and are available in [repository of choice, such as GBIF or a Symbiota portal] at [insert dataset DOI]."

IMPLEMENTING *THE WILD BEE DATA STANDARD*. Implementing a data standard requires adequate resources and infrastructure to ensure its credibility and longevity. Here, we have provided guidance on publishing data to GBIF, any insect-focused Symbiota portal, and iNaturalist (Fig. 1). Maintaining these portals demands dedicated personnel and funding to support ongoing user assistance, including data mapping, quality assessments, digitization efforts, and helping users access shared data. Additionally, using these portals requires adopting Darwin Core terminology to describe wild bee occurrence data. *The Wild Bee Data Standard*, which we introduce in the remainder of this manuscript, facilitates this transition by removing barriers for scientists who are unsure how to map or translate their data to align with Darwin Core. *The Wild Bee Data Standard* can be found on GitHub (https://github.com/Big-Bee-Network/wild-bee-data-standard), where comments and updates can be submitted and incorporated. Full implementation of *The Wild Bee Data Standard* will require creating a formal working group to develop a code of conduct for using data following the standard, monitor the GitHub repository, and update the standard over time to ensure it remains in active development and evolves alongside changes in wild bee monitoring practices (see https://www.tdwg.org/community/dwc/).

As more federal and funding agencies begin to require openly accessible data from supported projects (*e.g.*, USDA, NSF; Nelson, 2022), implementing *The Wild Bee Data Standard* and following our proposed digitization pipeline provides a pathway to meeting these requirements. Given the importance of openly accessible data to pollinator conservation (Chesshire *et al*., 2023; Rousseau *et al*., 2024), we hope to see these practices widely adopted and supported in perpetuity.

THE WILD BEE DATA STANDARD, VERSION 1.0.0: GUIDANCE FOR USE

The Wild Bee Data Standard consists of metadata guidelines and a list of 75 Darwin Core terms, or data fields, that describe wild bee occurrence data (Appendix 1). All 75 terms do not need to be used; the term list is divided into 26 *core* terms, 22 *recommended*

Table 1. Organizational structure of *The Wild Bee Data Standard*. *The Wild Bee Data Standard* contains 75 Darwin Core terms within three tiers of data collection and reporting (Levenson *et al*., 2025b). Within these tiers, the terms are presented in Darwin Core categorical order. For more information on these categories and the complete list of Darwin Core terms, please visit https://dwc.tdwg.org/terms/

Tiers of data collection and reporting		
Core	Recommended	Optional
These 26 terms must be collected and provided.	These 22 terms should be provided if the information was collected.	These 27 terms can be provided if the information was collected, depending on one's objective(s).
Categories of Darwin Core terms presented in The Wild Bee Data Standard		
Record-level	Generic data that might apply to any type of record in a dataset.	
Occurrence	Data that describe the existence of an organism at a particular place at a particular time.	
MaterialEntity	Data that describe an entity that can be identified, exists for some period of time, and consists in whole or in part of physical matter while it exists.	
Event	Data that describe an action that occurs at some location during some time.	
Location	Data that describe a spatial region or named place.	
Identification	Data that describe a taxonomic determination.	
Taxon	Data that describe a group of organisms.	

terms, and 27 *optional* terms (Table 1). *Core* terms are those that we argue must be collected and provided with all current and future wild bee occurrence data to ensure data quality; these are listed first, followed by *recommended* terms that should be provided if collected and *optional* terms that can be provided if collected (Levenson *et al*., 2025b). The standard is intended to guide field collection and specimen processing procedures (including taxonomic verification), streamline digitization, and support FAIR data principles. We advise establishing field collection and specimen processing procedures with metadata guidelines and *core* terms in mind, then consider the *recommended* and *optional* terms and choose to record those relevant to experimental goals. Field-collected data can then be digitized using one of the provided templates (Du Clos *et al*., 2025). These digitized data are then ready for taxonomic verification and ultimately to be made openly available by uploading to GBIF, a Symbiota portal, or another web-based data aggregator, portal or repository. Data downloaded from these sources will then be fully interoperable with each other.

Creating metadata. When uploading occurrence data to GBIF, Symbiota, or another web-based data aggregator, portal, or repository, descriptive information about the dataset, called metadata, must also be provided. Metadata offers important source information for occurrence datasets, promoting interpretation and reusability by data users. To meet GBIF's minimum metadata requirements, please provide the required metadata fields listed in Table 2 when publishing wild bee occurrence data. Required metadata fields contain basic information about an occurrence dataset, including the title, description, and contact information. Recommended metadata fields include data collection methods and the proper citation for use when working with a dataset downloaded from GBIF. Providing highly descriptive metadata improves data quality and contributes to more rigorous analyses (Bloom *et al.*, 2021). When a new dataset is published to GBIF, the IPT translates the provided metadata into an EML-compliant

Table 2. Metadata fields required or recommended by GBIF for any published dataset.

Field name	Definition	Required or recommended
Title	The title of your dataset. Aim to provide a detailed title that will distinguish your dataset from others.	Required
Description	A brief description of the dataset that may include where, when, and why the data were collected, who collected it, and what research conclusions the data has provided.	Required
Publishing organization	The organization publishing the dataset to GBIF. Organizations that have published wild bee occurrence data on GBIF include the USGS Bee Lab at the Eastern Ecological Science Center (previously the Bee Inventory and Monitoring Laboratory, Droege & Maffei, 2025) and the USDA-ARS Pollinating Insects Research Unit (Ikerd, 2019; Carril *et al*., 2023).	Required
Type	Choose from Occurrence, Taxon, or Event. Most data uploaded following The Wild Bee Data Standard will be of the Occurrence type.	Required
License	Choose a Creative Commons License (https://creativecommons.org/share-your-work/cclicenses/) to apply to your dataset. GBIF encourages publishers to adopt the least restrictive license possible from among three provided options: CC0 1.0, CC-BY 4.0, or CC-BY-NC 4.0.	Required
Contact(s)	Name and contact information for those with knowledge of the dataset.	Required
Creator(s)	Name and contact information for those who created the dataset, in priority order.	Required
Metadata provider(s)	Name and contact information for those who created the metadata for the dataset, in priority order.	Required
Sampling methodology	Providing information about the study extent in space and time and a description of sampling protocol and effort promotes transparency and reproducibility. More details are provided in the Metadata section of the IPT User Manual: https://ipt.gbif.org/manual/en/ipt/latest/manage-resources#sampling-methods	Recommended
Citation	Please provide a citation to use when citing the dataset. This citation can link to the dataset page on GBIF or it can link to any scientific publication that uses the provided data. More details are provided in the Metadata section of the IPT User Manual: https://ipt.gbif.org/manual/en/ipt/latest/manage-resources#citations	Recommended

metadata document, similar to how it maps Darwin Core terms (Robertson *et al*., 2014; Jones *et al*., 2019). For more information, please see the Metadata section of the IPT User Manual: https://ipt.gbif.org/manual/en/ipt/latest/manage-resources#metadata. We provide a GBIF-compliant metadata template with our worksheet and workbook data standard templates (Du Clos *et al*., 2025), which contains the terms listed in Table 2. See Droege & Maffei (2025) for an example of highly descriptive metadata for a wild bee occurrence dataset on GBIF.

Publishing metadata to a Symbiota portal is similar, though the upload to Symbiota will ask for slightly different metadata fields (https://docs.symbiota.org/Collection_Manager_Guide/editing_collection_metadata/#collections-metadata). Importantly, if your dataset is published to GBIF through a Symbiota portal, the relevant metadata provided to Symbiota will also be shared to GBIF, meeting their minimum metadata requirements. If using another CMS to publish to GBIF or publishing to another data

aggregator, portal, or repository, please determine their metadata requirements and prepare accordingly before publication.

Generating unique identifiers. Unique identifiers simplify and standardize data management and sharing practices. In *The Wild Bee Data Standard*, we describe two types of unique identifiers. Machine-readable identifiers are typically computer-generated, long, complex, and ideally globally unique. They are intended for use by software when sharing, managing, or analyzing data. An example of a machine-readable identifier is: 1edef0b2-df7e-4e0b-ab8a-6c367e622206. Human-usable identifiers are created and used by individuals and organizations managing data but are not expected to be unique outside of a given dataset or organization. Examples include specimen label numbers, site numbers, or combinations of site, date, and location information created to distinguish sampling events. Darwin Core provides fields to accommodate both machine-readable and human-usable identifiers.

There is one machine-readable identifier in *The Wild Bee Data Standard*: **dwc:occurrenceID**. In order to upload data to GBIF, **dwc:occurrenceID**s must be provided for each record in a dataset. While **dwc:occurrenceID**s can be created automatically through the use of a CMS to manage and share wild bee occurrence data, which we elaborate on below, they will sometimes need to be created by data providers. Creating unique **dwc:occurrenceID**s for each record is a critically important task. We recognize that this can present a barrier to preparing occurrence data for upload to GBIF and provide the following guidance. While it is acceptable for **dwc:occurrenceID**s to be unique within a dataset, we advise making them globally unique identifiers. Creating **dwc:occurrenceID**s has often been done by combining various cells, including event, location, and taxon identifiers, into an identifier unique within the dataset, but an issue with this approach is that even these resulting identifiers may not be globally unique. As such, we advise creating universally unique identifiers (UUIDs) to serve as **dwc:occurrenceID**s. UUIDs are 128-bit labels and can be created via R functions, Excel macros, or website applications, examples of which are provided in Appendix 3. Importantly, most CMS will create UUID-based **dwc:occurrenceID**s for each record created or uploaded, eliminating the need for data providers to create these identifiers on their own. For example, datasets uploaded to a Symbiota portal will have UUID-based **dwc:occurrenceID**s created for each record automatically. Using Symbiota or another CMS to manage occurrence data and share it to GBIF presents an approachable solution to generating UUIDs to serve as **dwc:occurrenceID**s for each record in an occurrence dataset. We reiterate that unique **dwc:occurrenceID**s for each record in an occurrence dataset are required for publication to GBIF.

Remember that **dwc:occurrenceID**s are long and complex for machine readability, not human usability. GBIF and other data management and analysis systems use **dwc:occurrenceID**s to keep track of each record in an occurrence dataset. When a dataset is modified and reuploaded, the **dwc:occurrenceID**s can be matched to update the modified records. GBIF has practices in place to monitor changes in **dwc:occurrenceID**s and notifies publishers if it detects new **dwc:occurrenceID**s for existing datasets (https://data-blog.gbif.org/post/improve-identifier-stability/). Managing data through Symbiota or other CMS makes modification of **dwc:occurrenceID**s extremely difficult, generally avoiding the same record ending up with multiple **dwc:occurrenceID**s. To prepare downloaded data for analyses, **dwc:occurrenceID**s can be used to filter out duplicates in cases where data may have been downloaded from multiple sources, such as GBIF or a Symbiota portal.

Human-usable identifiers are important in identifying a specimen or field site within a dataset, as in the terms **dwc:catalogNumber** and **dwc:fieldNumber**, which identify a specimen in a collection and a field site where occurrence data were collected. We describe the use of other human-usable identifiers in *The Wild Bee Data Standard* to help organize occurrence data internally within spreadsheet workbooks or relational databases. These identifiers include **dwc:eventID** to identify a unique sampling event, **dwc:locationID** to identify a unique sampling site location, and **dwc:taxonID** to identify a unique taxon name. The workbook template provided with *The Wild Bee Data Standard* uses these identifiers to separate sampling event information, site location information, and species identification information from the occurrence information of the wild bee specimen or observation (Du Clos *et al*., 2025). Providing the human-usable identifiers in the Darwin Core archive when uploading could aid data users in interpretation of the dataset and provide another way to remove duplicate records when preparing downloaded data for analyses.

Other unique identifiers mentioned in *The Wild Bee Data Standard* are connected to the places or people associated with the occurrence data. The IDs used in the terms **dwc:institutionID** and **dwc:collectionID** are generated by people working at the institution housing the data or the specimen collection. These can be cataloged in the Global Registry of Scientific Collections (https://www.gbif.org/grscicoll), though they can be other identifiers. The IDs used in the terms **dwc:recordedByID** and **dwc:identifiedByID** are created by the people doing the data collection and identification. Typically, these will be ORCID iDs, though they can be other unique identifiers. We recommend these terms to connect and properly credit relevant places and personnel with hosted or handled occurrence data.

USING CONTROLLED VOCABULARIES. Some of the terms included in *The Wild Bee Data Standard* have controlled vocabularies, or an allowable list of entries. Controlled vocabularies help standardize large amounts of data derived from different sources by allowing all users to glean the same information across multiple datasets. This makes the data more interoperable. Terms in *The Wild Bee Data Standard* with controlled vocabularies include: **dwc:basisOfRecord**, **dwc:occurrenceStatus**, **dwc:samplingProtocol**, **dwc:sampleSizeUnit**, **dwc:stateProvince**, **dwc:disposition**, **dwc:identificationQualifier**, **dwc:sex**, **dwc:caste**, **dwc:lifeStage**, **dwc:vitality**, **dwc:behavior**, **dwc:habitat**, **dwc:organismQuantityType**, **dwc:associatedTaxa**, and **dwc:dynamicProperties**. Further, **dwc:samplingProtocol**, **dwc:associatedTaxa**, **dwc:associatedOccurrences**, **dwc:georeferenceRemarks,** **dwc:dynamicProperties**, and **dwc:habitat** are or can be reported using key:value pairs to format the complex information within. In a key:value pair, the key is the type of information being provided, and the value is the actual information. Multiple key:value pairs can be reported in a single term. All terms with controlled vocabularies include a controlled vocabulary list of all allowable entries, and when applicable, the use of key:value pairs is thoroughly described for relevant terms.

FOLLOWING STANDARD TEMPLATES. To aid in the adoption and publication of data following *The Wild Bee Data Standard*, we provide two templates for entering compliant data into a computer (10.5281/zenodo.14187861; Du Clos *et al*., 2025). Both templates contain descriptive information and instructions for use in a README tab, a metadata tab that can be used to populate metadata when publishing online, and a full-term list

with definitions and links to their Darwin Core webpage. Data following the Darwin Core standard are presented in a single worksheet with each DwC term listed as a column heading. Therefore, we provide a single worksheet template that demonstrates the use of all terms in *The Wild Bee Data Standard*, with *core* terms in bold. At minimum, all *core* terms must be provided. The template can be modified as needed to provide additional data.

We also provide a workbook template as an option for those who use multiple spreadsheets in a workbook or a relational database to manage their data. We urge caution with the workbook template and only advise using it if your data management scheme follows its format. The workbook template represents one of multiple possible architectures for a database of wild bee occurrence data, a many-to-one database with occurrence information as the primary table. The workbook template contains an occurrence worksheet, an event worksheet, a location worksheet, and a taxon worksheet. All relevant terms are provided within their respective worksheets, with *core* terms in bold. At minimum, all *core* terms must be provided. The workbook template uses three human-readable unique identifiers to connect the worksheets together: **dwc:eventID**, **dwc:locationID**, and **dwc:taxonID**. Since uploading to web portals does not support multiple tables, the workbook template then concatenates all sheets to generate a large single worksheet suitable for uploading to GBIF, Symbiota, or another web-based data aggregator, portal, or repository.

We provide example data entries in both templates that describe how data are recorded for the case studies presented in the protocols developed by the Bee Monitoring RCN (Cariveau *et al*., 2025; Levenson *et al*., 2025a; López-Uribe *et al*., 2025; Otto *et al*., 2025; Strange *et al*., 2025), and provide scenarios here.

EXAMPLES OF IMPLEMENTING *THE WILD BEE DATA STANDARD*

Collecting community-level bee data examples. To demonstrate how the standard can be used to collect community level bee data, three case studies are presented in Levenson *et al*. (2025a). Our worksheet and workbook template examples (Du Clos *et al*., 2025) report data collected in the following scenarios, with full details for each provided in Levenson *et al*. (2025a). The following case studies describe several different bee data collection scenarios and the associated Darwin Core terms that would be used to record each piece of data:

Case Study I. An Inventory

Two entries are provided here, one for a bee caught in a bowl trap and one for a bee caught with a hand net. The first entry describes a bee collected by Field Scientist (**dwc:recordedBy**) in a bowl trap (**dwc:samplingProtocol**) at site 12 (**dwc:fieldNumber**, **dwc:locationID**) on July 12, 2009 (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**) at 4:00 pm (**dwc:eventTime**). The bee is later identified as *Agapostemon texanus* Cresson, 1872 (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**, **dwc:scientificNameAuthorship**) by Bee Taxonomist (**dwc:identifiedBy**). Further, Bee Taxonomist provided citations for the literature used to make this identification (**dwc:namePublishedIn**, **dwc:identificationReferences**). The second entry describes a bee collected by Field Scientist (**dwc:recordedBy**) in a net (**dwc:samplingProtocol**) while foraging on *Rubus parviflorus* (**dwc:associatedTaxa**) at site 6 (**dwc:fieldNumber**, **dwc:locationID**) on August 23, 2009 (**dwc:eventDate**, **dwc:year**, **dwc:month**,

dwc:day) at 1:20 pm (**dwc:eventTime**). This bee is later identified as *Bombus nevadensis* Cresson, 1874 (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**, **dwc:scientificNameAuthorship**) by Bee Taxonomist (**dwc:identifiedBy**). Further, Bee Taxonomist provided citations for the literature used to make this identification (**dwc:namePublishedIn**, **dwc:identificationReferences**). For both entries, site coordinates are taken from Google Maps (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:georeferenceRemarks**) and reported to six decimal places (**dwc:coordinateUncertaintyInMeters**, **dwc:coordinatePrecision**). The researchers share their data under public domain (**dwc:license**, **dwc:rightsHolder**).

Case Study II. A Survey

Two entries are provided here, one for a bee caught in a bowl trap and one for a bee caught with a hand net. The first entry describes a bee collected by Field Scientist (**dwc:recordedBy**) in a bowl trap (**dwc:samplingProtocol**) at site 8A (**dwc:fieldNumber**, **dwc:locationID**) on May 6, 2017 (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**) after a 24-hour deployment (**dwc:eventDate**). The bee is later identified as *Lasioglossum* cf. *versatum* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**, **dwc:verbatimIdentification**, **dwc:identificationQualifier**, **dwc:taxonRank**) by Bee Taxonomist (**dwc:identifiedBy**). Further, Bee Taxonomist provided a citation for the nomenclatural revision used to make this identification (**dwc:nameAccordingTo**). The second entry describes a bee collected by Field Scientist (**dwc:recordedBy**) in a net (**dwc:samplingProtocol**) while foraging on *Monarda fistulosa* (**dwc:associatedTaxa**) at site 8A (**dwc:fieldNumber**, **dwc:locationID**) on May 6, 2017 (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**) between 10:40 and 10:50 am (**dwc:eventTime**). This bee is later identified as *Osmia lignaria* subsp. *lignaria* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**, **dwc:infraspecificEpithet**) by Bee Taxonomist (**dwc:identifiedBy**). For both entries, site coordinates are taken from Google Maps (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:georeferenceRemarks**) and reported to six decimal places (**dwc:coordinateUncertaintyInMeters**, **dwc:coordinatePrecision**). The researchers share their data under a CC-BY license (**dwc:license**, **dwc:rightsHolder**).

Case Study III. A Monitoring Program

Three entries are provided here for three specimens of the same species collected by Field Scientist (**dwc:recordedBy**) in a bowl trap (**dwc:samplingProtocol**) at site 4 (**dwc:fieldNumber**, **dwc:locationID**) on September 13, 2021 (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**) after a 24-hour deployment (**dwc:eventDate**). These bees are all later identified as *Megachile relativa* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**) by Bee Taxonomist (**dwc:identifiedBy**). For all entries, site coordinates are taken from a GPS unit and reported to four decimal places (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:coordinateUncertaintyInMeters**, **dwc:coordinatePrecision**). The researchers share their data under public domain (**dwc:license**, **dwc:rightsHolder**).

OCCUPANCY OF FOCAL BEE SPECIES EXAMPLE. To demonstrate how the standard can be used in monitoring for occupancy of focal species, we provide an example aligned with the *Bombus affinis* monitoring framework described in Otto *et al*. (2025). Steps 3–5 of this framework are directly applicable to the use of *The Wild Bee Data Standard*. Our

worksheet and workbook template examples (Du Clos *et al*., 2025) report details of the sampling scheme created in Step 3 of the framework, including sampling unit size (**dwc:samplingEffort**), date sampled (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**), time of day sampled (**dwc:eventTime**), amount of time sampled (**dwc:samplingEffort**), observer experience (**dwc:eventRemarks**), land cover type sampled (**dwc:habitat**), caste (**dwc:caste**), and photo voucher metadata (**dwc:associatedMedia**). The template examples also report details associated with the field sampling methods described in Step 4 of the framework, including detections and non-detections (**dwc:occurrenceStatus**) and weather conditions (**dwc:dynamicProperties**). The template examples report data collected in the following scenario, which represents one collection event:

Three sites (**dwc:fieldNumber**, **dwc:locationID**) in the St. Paul, Minnesota area are sampled in early August 2023 (**dwc:eventDate**, **dwc:year**, **dwc:month**, **dwc:day**). The exact coordinates for these sites are masked, as *B. affinis* is an endangered species (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:informationWithheld**). As determined by remotely sensed data, two sites are in urban areas, and one is in grassland (**dwc:habitat**). Visual sampling (**dwc:samplingProtocol**) is conducted in the daytime for 30 minutes by a single observer, reported in person-hours (**dwc:eventTime**, **dwc:samplingEffort**). Weather conditions are reported before each sampling event (**dwc:dynamicProperties**). Data are recorded for each sampling event even if *B. affinis* was not detected (**dwc:eventID**, **dwc:occurrenceID**). *B. affinis* is detected at one site and is not detected at two sites (**dwc:occurrenceStatus**, **dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**). At the sites where *B. affinis* was not detected, **dwc:individualCount** is reported as zero. At the site where *B. affinis* was detected, three workers (**dwc:individualCount**, **dwc:caste**) were observed foraging on *Spiraea alba* (**dwc:associatedTaxa**) and one drone (**dwc:individualCount**, **dwc:caste**) was observed foraging on *Eutrochium purpureum* (**dwc:associatedTaxa**). Two photo vouchers are captured; one of a worker foraging on *S. alba* and one of the drone foraging on *E. purpureum*. URLs for both vouchers are shared on iNaturalist (**dwc:associatedMedia**). The researchers share their observation data in the public domain (**dwc:license**).

To comply with the framework, these surveys must be repeated at regular intervals throughout a defined sampling season. Data collection and recording began for this season on June 15, and will continue until September 15. The goal is to complete six 30-minute surveys at each sampling unit over the sampling season. Ideally, surveys take place during multiple sampling unit visits, spaced roughly two weeks apart. A single sampling visit may include one or multiple observers and each observer should be treated as an independent survey. When multiple observers are used, survey information for each observer should be recorded separately. All preceding and subsequent surveys will be recorded and reported as demonstrated here.

Collecting data on bee-flower interaction examples. To demonstrate how the standard can be used to collect bee-flower interaction data, two case studies are presented in Cariveau *et al*. (2025). Our worksheet and workbook template examples (Du Clos *et al*., 2025) report data collected in the following scenarios, with full details for each provided in Cariveau *et al*. (2025):

Case Study I. An Inventory for a Government Agency

Four entries are provided here, two for non-bumble bees that were lethally collected and two for bumble bees that were captured, cooled, photographed, and released.

The bees were sampled at Site 6 (**dwc:fieldNumber**), a dry prairie (**dwc:habitat**), on July 27, 2023 (**dwc:eventDate**), during the morning sampling period of the inventory (**dwc:eventID**). There is a single collector, Agency Technician, (**dwc:recordedBy**, **dwc:sampleSizeValue**, **dwc:SampleSizeUnit**) walking the 400 x 2 m meandering transect (**dwc:samplingEffort**) for a 1-hour sampling duration (**dwc:samplingEffort**). For each bee collected, the plant species it was caught on and the nature of the interaction is recorded (**dwc:associatedTaxa**). The timer is stopped when handling bees, therefore the total event time is reported as 1 hour and 50 minutes (**dwc:EventTime**). One non-bumble bee was collected on a plant that was later photographed as the representative voucher for that plant species (**dwc:associatedOccurrences**). The time the bumble bees spent in the cooler before being photographed was recorded (**dwc:preparations**). Site coordinates were obtained with a handheld GPS (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:georeferenceRemarks**) and reported to four decimal places, for an accuracy of 5 m (**dwc:coordinateUncertaintyInMeters**). The non-bumble bees were later identified as *Lasioglossum subviridatum* and *Melissodes bimaculatus* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**) by Bee Taxonomist (**dwc:identifiedBy**). The specimens were then accessioned at Local Museum (**dwc:collectionCode**). Specimen data were reported under a CC-BY license (**dwc:license**). The bumble bee photo vouchers were posted to iNaturalist (**dwc:associatedMedia**). They were identified in the field by Agency Technician (**dwc:identifiedBy**) as *Bombus ternarius* and *Bombus bimaculatus* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**). The bee photo vouchers were shared under public domain (**dwc:license**).

Case Study II. Monitoring to Determine Effects of Prescribed Fire

Three entries are provided here, two from a burned site and one from an unburned site. These two sites (**dwc:fieldNumber**) were sampled on the same day (May 21, 2019; **dwc:eventDate**) during the afternoon sampling period (**dwc:eventTime**), but by different teams of two technicians (**dwc:sampleSizeValue**, **dwc:sampleSizeUnit**, **dwc:recordedBy**). At each site, the technicians walked along a 200 m meandering transect, collecting bees on 1 m of either side of the center of the transect (**dwc:samplingEffort**), for a sampling duration of 10 minutes, reported in decimals of person-hours (**dwc:samplingEffort**). The timer is stopped when handling bees, so **dwc:eventTime** is longer than 30 minutes. For each bee collected, the plant species it was caught on and the nature of the interaction was recorded (**dwc:associatedTaxa**). One bee was collected on a plant that was later photographed as the representative voucher for that plant species (**dwc:associatedOccurrences**). Site coordinates were confirmed in Google Maps (**dwc:decimalLatitude**, **dwc:decimalLongitude**, **dwc:georeferenceRemarks**) and reported to six decimal places, for an accuracy of 0.1 m (**dwc:coordinateUncertaintyInMeters**). The bees at the burned site were later identified as *Lasioglossum abanci* and *Andrena imitatrix* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**) by Bee Taxonomist (**dwc:identifiedBy**). The bee collected at the burned site was identified in the field by Student Technician (**dwc:identifiedBy**) as *Bombus impatiens* (**dwc:scientificName**, **dwc:genus**, **dwc:specificEpithet**). The specimens were accessioned into the host university's invertebrate zoology collection (**dwc:collectionCode**). The data were shared under public domain (**dwc:license**). This scenario was developed using locations, habitats, and species found in Ulyshen *et al*. (2022).

ACKNOWLEDGEMENTS

We thank all 800+ members of the U.S. National Native Bee Monitoring Research Coordination Network. We are grateful to all presenters and participants who attended the Bee Monitoring RCN Data Management Workshop in March 2023. Izzy Hill co-organized the Bee Monitoring RCN Data Management Workshop with authors B. Du Clos, K.C. Seltmann, E.M. Tucker, and S.H. Woodard and provided critical feedback on early versions of the standard. E.M. Tucker was supported in this work by NSF award DBI-2216927. K.C. Seltmann was supported in this work by NSF Award DBI-2102006. Ryan Drum and Kristen Dillon facilitated the connection between this standard and the BOMS standard of the U.S. Fish and Wildlife Service. The group of experts who developed the RCN protocols provided critical feedback on a later draft of the standard. We thank reviewers Abigail Benson, David Bloom, Paige Chesshire, James Dorey, and Josee Rousseau and *Journal of Melittology* editors Victor Gonzalez and Claus Rasmussen for their comments and suggestions, which substantially improved the manuscript. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service. This work was supported by the Pollinator Health Program, project award no. 2020-67014-38165 [awarded to S.H.W.], from the U.S. Department of Agriculture's National Institute of Food and Agriculture.

REFERENCES

A Collective and Ongoing Collaborative Effort by Those Who Love to Study Bees in North America. 2024. The Very Handy Bee Manual (2.0). Zenodo. [https://doi.org/10.5281/zenodo.12812755; last accessed October 2024.]

Bartomeus, I., J.R. Stavert, D. Ward, & O. Aguado. 2019. Historical collections as a tool for assessing the global pollination crisis. *Philosophical Transactions of the Royal Society B* 374(1763): 20170389. https://doi.org/10.1098/rstb.2017.0389

Bloom D.A., P. Zermoglio, & R. Guralnick. 2021. Analysis of biodiversity data needs in the post-2020 framework. Copenhagen: GBIF Secretariat. https://doi.org/10.35035/doc-2ph8-0403

Boone, M.L., Z.M. Portman, I. Lane, & S. Rao. 2023. Occupancy of *Bombus affinis* (Hymenoptera: Apidae) in Minnesota is highest in developed areas when standardized surveys are employed. *Environmental Entomology* 52(5): 918–938. https://doi.org/10.1093/ee/nvad088

Brand, A., L. Allen, M. Altman, M. Hlava, & J. Scott. 2015. Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing* 28(2): 151–155. https://doi.org/10.1087/20150211

Briggs, E.L., C. Baranski, O. Münzer Schaetz, G. Garrison, J.A. Collazo, & E. Youngsteadt. 2022. Estimating bee abundance: can mark-recapture methods validate common sampling protocols? *Apidologie* 53(10). https://doi.org/10.1007/s13592-022-00919-4

Burkle, L.A., J.C. Knight, & T.M. Knight. 2013. Plant-pollinator interactions over 120 years: Loss of species, co-occurrence, and function. *Science* 339(6127): 1611–1615. https://doi.org/10.1126/science.1232728

Burrows, S., C. Ritner, M. Christman, L. Spears, A. Smith-Pardo, S. Price, R. Ramirez, T. Griswold, & A. Redford. 2021. Exotic Bee ID: About bees. [https://idtools.org/exotic_bee/index.cfm?pageID=3064; accessed 14 June 2024.]

Campbell, C.J., V. Barve, M.W. Belitz, J.R. Doby, E. White, C. Seltzer, G. Di Cecco, A.H. Hurlbert, & R. Guralnick. 2023. Identifying the identifiers: How iNaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science. *BioScience* 73(7): 533–541. https://doi.org/10.1093/biosci/biad051

Cariveau, D.P., K.-L.J. Hung, N.M. Williams, D.W. Inouye, C.T. Burns, I.G. Lane, R.E. Irwin, H.K. Levenson, B. Du Clos, & S.H. Woodard. 2025. Standardized protocols for collecting data on bee-flower interactions and the associated floral community. *Journal of Melittology* 123(5): 104–138. https://doi.org/10.17161/jom.vi123.23861

Carril, O., T. Griswold, & H.I. Ikerd. 2023. Wild bees of Grand Staircase-Escalante National Monument: Richness, abundance, and spatio-temporal beta-diversity. PeerJ 6: e5867. https://doi.org/10.7717/peerj.5867 [Occurrence dataset accessed via GBIF.org on 2024-11-04]

Carroll, S.R., I. Garba, O.L. Figueroa-Rodríguez, J. Holbrook, R. Lovette, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J.D. Walker, J. Anderson, & M. Hudson. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19(43): 1–12. https://doi.org/10.5334/dsj-2020-043.

Chesshire, P. 2023. Full Dataset prior to Cleaning. figshare. Dataset. https://doi.org/10.6084/m9.figshare.22455616.v1

Chesshire, P.R., E.E. Fischer, N.J. Dowdy, T.L. Griswold, A.C. Hughes, M.C. Orr, J.S. Ascher, L.M. Guzman, K.L.J. Hung, N.A. Cobb, & L.M. McCabe. 2023. Completeness analysis for over 3000 United States bee species identifies persistent data gap. *Ecography* 2023(5): e06584. https://doi.org/10.1111/ecog.06584

Cicero, C., M.S. Koo, E. Braker, J. Abbott, D. Bloom, M. Campbell, J.A. Cook, J.R. Demboski, A.C. Doll, L.M. Frederick, A.J. Linn, T.J. Mayfield-Meyer, D.L. McDonald, M.W. Nachman, L.E. Olson, D. Roberts, D.S. Sikes, C.C. Witt, & E.A. Wommack. 2024. Arctos: Community-driven innovations for managing natural and cultural history collections. *PLoS ONE* 19(5): e0296478. https://doi.org/10.1371/journal.pone.0296478

Cobb, N.S., L.F. Gall, J.M. Zaspel, N.J. Dowdy, L.M. McCabe, & A.Y. Kawahara. 2019. Assessment of North American arthropod collections: Prospects and challenges for addressing biodiversity research. *PeerJ* 7: e8086 https://doi.org/10.7717/peerj.8086

Dorey, J.B., E.E. Fischer, P.R. Chesshire, A. Nava-Bolaños, R.L. O'Reilly, S. Bossert, S.M. Collins, E.M. Lichtenberg, E.M. Tucker, A. Smith-Pardo, A. Falcon-Brindis, D.A. Guevara, B. Ribeiro, D. de Pedro, J. Pickering, K.L.J. Hung, K.A. Parys, L.M. McCabe, M.S. Rogan, R.L. Minckley, S.J.E. Velazco, T. Griswold, T.A. Zarrillo, W. Jetz, Y.V. Sica, M.C. Orr, L.M. Guzman, J.S. Ascher, A.C. Hughes, & N.S. Cobb. 2023. A globally synthesised and flagged bee occurrence dataset and cleaning workflow. *Scientific Data* 10(747). https://doi.org/10.1038/s41597-023-02626-w

Droege. S., & C. Maffei. 2025. Insect species occurrence data from multiple projects worldwide with focus on bees and wasps in North America. Version 1.12. United States Geological Survey. Sampling event dataset. https://doi.org/10.15468/6autvb accessed via GBIF.org on 2025-02-14.

Du Clos, B., K.C. Seltmann, N.E. Turley, C. Maffei, E.M. Tucker, I.G. Lane, H.K. Levenson, & S.H. Woodard. 2025. Templates for The Wild Bee Data Standard (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.14187861

Duke, C.S. & J.H. Porter. 2013. The ethics of data sharing and reuse in biology. *BioScience* 63(6): 483–489. https://doi.org/10.1525/bio.2013.63.6.10

GBIF.org (03 February 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.6cxfsw

Graham, K.K., J. Gibbs., J. Wilson, E. May, & R. Isaacs. 2021. Resampling of wild bees across fifteen years reveals variable species declines and recoveries after extreme weather. *Agriculture, Ecosystems & Environment* 317(107470). https://doi.org/10.1016/j.agee.2021.107470

Graves, T.A., W.M. Janousek, S.M. Gaulke, A.C. Nicholas, D.A. Keinath, C.M. Bell, S. Cannings, R.G. Hatfield, J.M. Heron, J.B. Koch, H.L. Loffland, L.L. Richardson, A.T. Rohde, J. Rykken, J.P. Strange, L.M. Tronstad, & C.S. Sheffield. 2020. Western bumble bee: Declines in the continental United States and range-wide information gaps. *Ecosphere* 11(6): e03141. https://doi.org/10.1002/ecs2.3141

Gries, C., E. Gilbert, & N. Franz. 2014. Symbiota – A virtual platform for creating voucher-based biodiversity information communities. *Biodiversity Data Journal* 2: e1114. https://doi.org/10.3897/BDJ.2.e1114

Haak, L.L., M. Fenner., L. Paglione, E. Pentz, & H. Ratner. 2012. ORCID: A system to uniquely identify researchers. *Learned Publishing* 25(4): 259–264. https://doi.org/10.1087/20120404

Jones, M.B., M. O'Brien, B. Mecum, C. Boettiger, M. Schildhauer, M. Maier, T. Whiteaker, S. Earl, & S. Chong. 2019. Ecological Metadata Language version 2.2.0. *KNB Data Repository*. https://doi.org/10.5063/F11834T2

Ikerd, H. 2019. Bee Biology and Systematics Laboratory. USDA-ARS Pollinating Insect-Biology, Management, Systematics Research. Occurrence dataset https://doi.org/10.15468/anyror accessed via GBIF.org on 2024-11-04.

Lendemer. J., B. Thiers, A.K. Monfils, J. Zaspel, E.R. Ellwood, A. Bentley, K. LeVan, J. Bates, D. Jennings, D. Contreras, L. Lagomarsino, P. Mabee, L.S. Ford, R. Guralnick, R.E. Gropp, M. Revelez, N. Cobb, K. Seltmann, & M.C. Aime. 2020. The Extended Specimen Network: A strategy to enhance US biodiversity collections, promote research and education. *BioScience* 70(1): 23–30. https://doi.org/10.1093/biosci/bizl140

Levenson, H.K., B.N. Metz, & D.R. Tarpy. 2024. Effects of study design parameters on estimates of bee abundance and richness in agroecosystems: A meta-analysis. *Annals of the Entomological Society of America* 117(2): 92–106. https://doi.org/10.1093/aesa/saae001

Levenson, H.K., O. Messinger Carril, N.E. Turley, C. Maffei, G. LeBuhn, T. Griswold, N.M. Williams, K.-L.J. Hung, R.E. Irwin, B. Du Clos, & S.H. Woodard. 2025a. Standardized protocol for collecting community-level bee data. *Journal of Melittology* 123(4): 78–103. https://doi.org/10.17161/jom.vi123.22649

Levenson, H.K., B. Du Clos, T.A., Smith, S. Jepsen, J.G. Everett, N.M. Williams, & S.H. Woodard. 2025b. A call for standardization in wild bee data collection and curation. *Journal of Melittology* 123(2): 4–16. https://doi.org/10.17161/jom.vi123.22533

López-Uribe, M.M., J.P. Strange, L. Whiteman, B.N. Danforth, S. Jha, M.G. Branstetter, J.B.U. Koch, H.K. Levenson, B. Du Clos, & S.H. Woodard. 2025. Standardized protocol for collecting bee samples to generate molecular data. *Journal of Melittology* 123(7): 163–181. https://doi.org/10.17161/jom.vi123.22596

MacPhail, V.J., R. Hatfield, & S.R. Colla. 2024. Bumble Bee Watch community science program increases scientific understanding of an important pollinator group across Canada and the USA. *PLoS ONE* 19(5): e0303335. https://doi.org/10.1371/journal.pone.0303335

Martín-Mora, E., S. Ellis, & L.M. Page. 2020. Use of web-based species occurrence information systems by academics and government professionals. *PLoS ONE* 15(7): e0236556. https://doi.org/10.1371/journal.pone.0236556.

Mathiasson, M.E., & S.M. Rehan. 2019. Status changes in the wild bees of northeastern North America over 125 years revealed through museum specimens. *Insect Conservation and Diversity* 12(4): 278–288. https://doi.org/10.1111/icad.12347

Meineke, E.K., T.J. Davies, B.H. Daru, & C.C. Davis. 2018. Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B*. 374(1763): 20170386. https://doi.org/10.1098/rstb.2017.0386

Montgomery, G.A., M.W. Belitz, R.P. Guralnick, & M.W. Tingley. 2021. Standards and best practices for monitoring and benchmarking insects. *Frontiers in Ecology and Evolution* 8: 579193. https://doi.org/10.3389/fevo.2020.579193

Nachman, M.W., E.J. Beckman, R.C.K. Bowie, C. Cicero, C.J. Conroy, R. Dudley, T.B. Hayes, M.S. Koo, E.A. Lacey, C.H. Martin, J.A. McGuire, J.L. Patton, C.L. Spencer, R.D. Tarvin, M.H. Wake, I.J. Wang, A. Achmadi, S.T. Álvarez-Castañeda, M.J. Anderson, J. Arroyave, C.C. Austin, F.K. Barker, L.N. Barrow, G.F. Barrowclough, J. Bates, A.M. Bauer, K.C. Bell, R.C. Bell, A.W. Bronson, R.M. Brown, F.T. Burbrink, K.J. Burns, C.D. Cadena, D.C. Cannatella, T.A. Castoe, P. Chakrabarty, J.P. Colella, J.A. Cook, J.L. Cracraft, D.R. Davis, A.R.D. Rabosky, G. D'Elìa, J.P. Dumbacher, J.L. Dunnum, S.V. Edwards, J.A. Esselstyn, J. Faivovish, J. Fjeldså, O.A. Flores-Villela, K. Ford, J. Fuchs, M.K. Fujita, J.M. Good, E. Greenbaum, H.W. Greene, S. Hackett, A. Hamidy, J. Hanken, T. Haryoko, M.T.R. Hawkins, L.R. Heaney, D.M. Hillis, B.D. Hollingsworth, A.D. Hornsby, P.A. Hosner, M. Irham, S. Jansa, R.A. Jiménez, L. Joseph, J.J. Kirchman, T.J. LaDuc, A.D. Leaché, E.P. Lessa, H. López-Fernández, N.A. Mason, J.E. McCormack, C.D. McMahan, R.G. Moyle, R.A. Ojeda, L.E. Olson, C.K. Onn, L.R. Parenti, G. Parra-Olea, B.D. Patterson, G.B. Pauly, S.E. Pavan, A.T. Peterson, S. Poe, D.L. Rabosky, C.J. Raxworthy, S. Reddy, A. Rico-Guevara, A. Riyanto, L.A. Rocha, S.R. Ron, S.M. Rovito, K.C. Rowe, J. Rowley, S. Ruane, D. Salazar-Valenzuela, A.J. Shultz, B. Sidlauskas, D.S. Sikes, N.B. Simmons, M.L.J. Stiassny, J.W. Streicher, B.L. Stuart, A.P. Summers, J. Tavera, P. Teta, C.W. Thompson, R.M. Timm, O. Torres-Carvajal, G. Voelker, R.S. Voss, K. Winker, C. Witt, E.A. Wommack, & R.M. Zink. 2023. Specimen collection is essential for modern science. *PLoS Biology*. 21(11): e3002318. https://doi.org/10.1371/journal.pbio.3002318

Nelson, A. 2022. Memorandum for the Heads of Executive Departments and Agencies: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. Executive Office of The President Office of Scientific Technology and Policy. 8 pp. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf. Accessed November 2024.

Nelson, G., & S. Ellis. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*. 374: 20170391. https://doi.org/10.1098/rstb.2017.0391

Orr, M.C., A.C. Hughes, D. Chesters, J. Pickering, C.D. Zhu, & J.S. Ascher. 2021. Global patterns and drivers of bee distribution. *Current Biology* 31(3): 451–458. https://doi.org/10.1016/j.cub.2020.10.053

Ostwald, M.M., V.H. Gonzalez, C. Chang, N. Vitale, M. Lucia, & K.C. Seltmann. 2024. Toward a functional trait approach to bee ecology. *Ecology and Evolution* 14(10): e70465. https://doi.org/10.1002/ece3.70465

Otto, C.R.V., L.L. Bailey, B. Du Clos, T. Smith, E. Evans, I. Pearse, S. Killingsworth, S. Jepsen, & S.H. Woodard. 2025. Estimating occupancy of focal bee species. *Journal of Melittology* 123(6): 139–162. https://doi.org/10.17161/jom.vi123.22555

Packer, L., Monckton, S.K., Onuferko, T.M., & R.R. Ferrari. 2018. Validating taxonomic identifications in entomological research. *Insect Conservation and Diversity* 11(1): 1–12. https://doi.org/10.1111/icad.12284

R Core Team. 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robertson, T., M. Döring, R. Guralnick, D. Bloom, J. Wieczorek, K. Braak, J. Otegui, L. Russell, & P. Desmet. 2014. The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the Internet. *PLoS ONE* 9(8): e102623. https://doi.org/10.1371/journal.pone.0102623

Rousseau, J.S., S.H. Woodard, S.J. Jepsen, B. Du Clos, A. Johnston, B.N. Danforth, & A.D. Rodewald. 2024. Advancing bee conservation in the US: Gaps and opportunities in data collection and reporting. *Frontiers in Ecology and Evolution* 12: 1346795. https://doi.org/10.3389/fevo.2024.1346795

Rousseau, J.S., A. Johnston, & A.D. Rodewald. 2025. Where the wild bees are: Birds improve indicators of bee richness. *Plos ONE* 20(4): e0321496. https://doi.org/10.1371/journal.pone.0321496

Seltmann, K., S. Lafia, D. Paul, S. James, D. Bloom, N. Rios, S. Ellis, U. Farrell, J. Utrup, M. Yost, E. Davis, R. Emery, G. Motz, J. Kimmig, V. Shirey, E. Sandall, D. Park, C. Tyrrell, R. Thackurdeen, M. Collins, V. O'Leary, H. Prestridge, C. Evelyn, & B. Nyberg. 2018. Georeferencing for Research Use (GRU): An integrated geospatial training paradigm for biocollections researchers and data providers. *Research Ideas and Outcomes* 4: e32449. https://doi.org/10.3897/rio.4.e32449

Seltmann, K. 2022. Natural history specimens collected and/or identified and deposited. Data set. Zenodo. https://doi.org/10.5281/zenodo.6575038

Shorthouse, D.P. 2020. Bionomia: Linking natural history specimens to the world's collectors. https://bionomia.net.

Sigovini, M., E. Keppel, & D. Tagliapietra. 2016. Open nomenclature in the biodiversity era. *Methods in Ecology and Evolution* 7: 1216–1225. https://doi.org/10.1111/2041-210X.12594

Strange, J.P., M.M. López-Uribe, L. Whiteman, B.N. Danforth, S. Jha, H.K. Levenson, B. Du Clos, J.B.U. Koch, & S.H. Woodard. 2025. Standardized protocol for collecting bee samples for internal parasite and pathogen data. *Journal of Melittology* 123(8): 182–194. https://doi.org/10.17161/jom.vi123.22598

Strasser, C., R. Cook, W. Michener. & A. Budden. 2012. Primer on data management: What you always wanted to know. *DataOne*. https://doi.org/doi:10.5060/D2251G48

Thessen, A.E., M. Woodbrun, D. Koureas, D. Paul, M. Conlon, D.P. Shorthouse, & S. Ramdeen. 2019. Proper attribution for curation and maintenance of research collections: Metadata recommendations of the RDA/TDWG working group. *Data Science Journal* 18:54. https://doi.org/10.5334/dsj-2019-054

Turley, N.E., S.E. Kania, I.R. Petitta, E.A. Otruba, D.J. Biddinger, T.M. Butzler, V.V. Sesler, & M.M. López-Uribe. 2024. Bee monitoring by community scientists: Comparing a collections-based program with iNaturalist. *Annals of the Entomological Society of America* 117(4): 220–233. https://doi.org/10.1093/aesa/saae014

Turney, S., E.R. Cameron, C.A. Cloutier, & C.M. Buddle. 2015. Non-repeatable science: Assessing the frequency of voucher specimen deposition reveals that most arthropod research cannot be verified. *PeerJ*: 3:e1168. https://doi.org/10.7717/peerj.1168

Ulyshen, M., K. Elliott, J. Scott, S. Horn, P. Clinton, N. Liu, C.F. Miniat, P. Caldwell, C. Oishi, J. Knoepp, & P. Bolstad. 2022. Effects of *Rhododendron* removal and prescribed fire on bees and plants in the southern Appalachians. *Ecology and Evolution* 12(3): e8677. https://doi.org/10.1002/ece3.8677

US Fish and Wildlife Service. 2023. Biotic Observation Minimum Specification for FWS Inventory and Monitoring Surveys. [https://ecos.fws.gov/ServCat/Reference/Profile/153885; last accessed 14 June 2024.]

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, & D. Vieglais. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson, M., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. de Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, W. Waagmeester, P. Wittenberg, K. Wolstencroft, J. Zhao, & B. Mons. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(160018). https://doi.org/10.1038/sdata.2016.18

APPENDIX 1
The Wild Bee Data Standard, version 1.0.0: List of Terms

This appendix presents the complete term list for *The Wild Bee Data Standard*, designed to ensure consistent collection and sharing of wild bee occurrence data. The standard includes 75 Darwin Core terms, organized into 26 core, 22 recommended, and 27 optional fields (see Table 1). Each entry provides a definition, rationale, usage instructions, and examples, with controlled vocabularies where appropriate. Links to the Darwin Core reference guide and alignment with Bee Monitoring RCN protocols are also provided. An example of the term entry format is as follows:

Term name: Term name in bold.
Definition: What information the term describes.
Rationale: Reason for using the term to describe wild bee occurrences.
Requirement: One of the following three categories (Table 1):
Core: These terms must be collected and provided.
Recommended: These terms should be provided if the information was collected.
Optional: These terms can be provided if the information was collected.
How to use: Instructions for using the term to align with *The Wild Bee Data Standard.*
Controlled vocabulary list: The list of allowable entries for terms with a controlled vocabulary.
Example(s): Practical example(s) of text that would be entered into a field for that term.
Relevant protocols: Most terms are relevant to all Bee Monitoring RCN protocols: Cariveau *et al*. (2025), Levenson *et al.* (2025a), López-Uribe *et al.*, (2025), Otto *et al.* (2025), and Strange *et al*. (2025).
For more detail: Link to the term entry in the Darwin Core quick reference guide (https://dwc.tdwg.org/terms/).

Term List Index

Core (26 terms)

Record-level: institutionCode, basisOfRecord, informationWithheld
Occurrence: occurrenceID, catalogNumber, recordedBy, individualCount, occurrenceStatus
Event: eventDate, eventTime, year, month, day, samplingProtocol, sampleSizeValue, sampleSizeUnit, samplingEffort
Location: country, stateProvince, decimalLatitude, decimalLongitude, coordinateUncertaintyInMeters
Identification: identifiedBy
Taxon: scientificName, genus, specificEpithet

Recommended (22 terms)

RECORD-LEVEL: license, institutionID, collectionID, collectionCode,
OCCURRENCE: recordedByID, sex, associatedTaxa
MATERIALENTITY: disposition
EVENT: eventID, eventRemarks
LOCATION: locationID, geodeticDatum
IDENTIFICATION: verbatimIdentification, identificationQualifier, identifiedByID, dateIdentified, identificationReferences
TAXON: taxonID, nameAccordingTo, family, infraspecificEpithet, scientificNameAuthorship

Optional (27 terms)

RECORD-LEVEL: rightsHolder, dynamicProperties
OCCURRENCE: lifeStage, caste, behavior, vitality, associatedMedia, associatedOccurrences, occurrenceRemarks
MATERIALENTITY: preparations, associatedSequences, materialEntityRemarks
EVENT: fieldNumber, habitat
LOCATION: county, locality, verbatimElevation, coordinatePrecision, georeferencedBy, georeferenceRemarks
IDENTIFICATION: typeStatus, identificationRemarks
TAXON: namePublishedIn, tribe, subgenus, taxonRank, vernacularName

Core Terms

Record-level

TERM NAME: **institutionCode**
DEFINITION: The name (or acronym) in use by the institution or organization having custody of the object(s) or information referred to in the record.
RATIONALE: Indicates the institution that holds or curates the specimen or observation record, which may lead to a point of contact for more information about an occurrence.
REQUIREMENT: Core.
HOW TO USE: Provide a name or acronym for the institution or organization that houses the specimen or coordinated the observation. If the institution or organization is registered in the Global Registry of Scientific Collections (https://scientific-collections.gbif.org/), please use the code listed there. For lab collections managed in-house, please provide the name of the lab.
EXAMPLES: USDA-ARS, UNHC, iNaturalist, AMNH, Woodard Lab
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:institutionCode

TERM NAME: **basisOfRecord**

Definition: The specific nature of the data record.
Rationale: This term is *required for publication to GBIF.* This term describes the type of occurrence record to indicate whether it includes a specimen, an observation, tissue material, or is gathered from the literature. This is essential context for interpretation.
Requirement: Core.
How to use: This term uses a controlled vocabulary. *The Wild Bee Data Standard* follows the GBIF vocabulary for the basis of record.

- Controlled vocabulary list for **dwc:basisOfRecord**:
  - PreservedSpecimen, to describe labeled wild bee specimens preserved on pins or in vials.
  - HumanObservation, to describe wild bees observed in the field and captured in photographs taken in person when using non-lethal sampling methods.
  - MachineObservation, to describe wild bees captured in photographs taken via camera trap when using non-lethal sampling methods.
  - MaterialEntity, taken from the Darwin Core vocabulary for **dwc:basisOfRecord** to describe genetic or pathogen material.
  - MachineCitation, to describe occurrences gathered from published literature.

Examples: PreservedSpecimen, HumanObservation
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:basisOfRecord
https://docs.gbif.org/course-data-use/en/basis-of-record.html

Term name: **informationWithheld**
Definition: Additional information that exists, but that has not been shared in the given record.
Rationale: If any core information is not shared, the type of information and the rationale for not sharing it must be provided. In particular, if the exact location of the record is masked, this cell must indicate what information has been omitted or changed, and describe why masking occurred.
Requirement: Core, if applicable (see rationale).
How to use: This is a text-based term that allows any word-based entry.
Examples:

- Exact location removed to mask for endangered/threatened/rare bee or host plant species or owing to private land ownership.
- Decimal latitude/longitude rounded to obscure exact location because it is private land.

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:informationWithheld

Occurrence

Term name: **occurrenceID**
Definition: A unique identifier for the occurrence.
Rationale: This term is *required for publication to GBIF.* Please see the section "Generating unique identifiers."

REQUIREMENT: Core.
HOW TO USE: We advise creating universally unique identifiers, or UUIDs, to serve as **dwc:occurrenceID**s. These can be created using R packages, Excel macros, or website applications, examples of which are provided in Appendix 3. While GBIF will not generate these for you, consider that uploading to a CMS automatically creates **dwc:occurrenceIDs** for each record. Please see the section "Generating unique identifiers" for more details.
EXAMPLE: 8d13f958-10fa-490a-8c32-5938be9d2037
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:occurrenceID
https://www.gbif.org/data-quality-requirements-occurrences#dcOccurrenceID

TERM NAME: **catalogNumber**
DEFINITION: A unique identifier within a dataset or collection.
RATIONALE: Physical specimens accessioned into a collection will have a unique ID assigned to them; that number gets reported through **dwc:catalogNumber** to reference back to the original dataset.
REQUIREMENT: Core, if applicable (see rationale).
HOW TO USE: Use **dwc:catalogNumber** to share the unique ID of a record or specimen in an original dataset or collection. To improve uniqueness, human-readability, and machine interpretation in large, multi-institution datasets, please include a sufficient amount of both letters and numbers when creating these IDs.
EXAMPLES: BBSL319283, AMNH PBI 82341, UNH-150861
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/list/#dwc_catalogNumber

TERM NAME: **recordedBy**
DEFINITION: A list of names of people, groups, or organizations responsible for recording the original occurrence. This is a synonym of "Collector."
RATIONALE: Providing the names of the collector or observer creates a potential point of contact for more information regarding an occurrence.
REQUIREMENT: Core.
HOW TO USE: Use full names (First [Middle Initial, if provided] Last) whenever possible. When multiple collectors are working simultaneously, list the specimens collected by each collector individually. If that is not possible, the primary collector or observer should be listed first. Separate the values in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell. Please be sure to record the number of collectors in **dwc:samplingEffort**.
EXAMPLES:

- For one collector:
  - Frank D. Parker
- For more than one collector:
  - Olivia J. Messinger | Terry L. Griswold

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:recordedBy

Term name: **individualCount**
Definition: The number of individual organisms present at the time of the occurrence.
Rationale: Most records of wild bee occurrences represent one bee. This indicates a detection of a bee, but it is important in many cases to also indicate non-detections of focal bee species, represented as zeros.
Requirement: Core.
How to use: Report the number of individual organisms counted in the occurrence. Only whole numbers are permitted. When sampling for occupancy of focal species (Otto *et al.*, 2025), if they are not found at a sampling site, use this term to report non-detections.

- A 0 entry for **dwc:individualCount** represents a non-detection.
- Most records will have an entry of 1 for **dwc:individualCount**, as most records correspond to an observation or collection (a detection) of a single organism.
- Entries larger than 1 are possible in some cases; for example:
  - If a large number of the same species is collected, but only a subset of specimens of that species are preserved.
  - If multiple bees of the same species are in a photo observation on a plant.
  - If multiple bees of the same species are observed at the same site, but a photo voucher is created for only one of those bees.

Examples: 0, 1, 2, etc.
Relevant protocols: All, but particularly important for the occupancy of focal species protocol (Otto *et al.*, 2025).
For more detail: https://dwc.tdwg.org/terms/#dwc:individualCount
https://www.gbif.org/data-quality-requirements-occurrences#dcCount

Term name: **occurrenceStatus**
Definition: A statement about the presence or absence of a taxon at a location.
Rationale: This term indicates if a species was detected or not detected during a sampling event and is particularly important to report when conducting occupancy surveys. Providing both detections and non-detections are crucial for occupancy modeling.
Requirement: Core.
How to use: This term uses a controlled vocabulary. Although the Darwin Core standard recommends the use of "present" and "absent" to designate occurrence status, we suggest observers instead use "detected" and "notDetected" when recording results from sampling events. Absence is seldom known with certainty and we recommend avoiding the use of this term during data collection.

- Controlled vocabulary list:
  - detected
  - notDetected

Examples: detected, notDetected
Relevant protocols: All, but particularly important for the occupancy of focal species protocol (Otto *et al.*, 2025).
For more detail: https://dwc.tdwg.org/terms/#dwc:occurrenceStatus

Event

TERM NAME: **eventDate**
DEFINITION: The date-time or interval during which a sampling event occurred.
RATIONALE: This term is *required for publication to GBIF*. Providing a collection or observation date provides essential context for interpreting an occurrence.
REQUIREMENT: Core.
HOW TO USE: Recommended best practice is to use a date that conforms to ISO 8601-1:2019. Provide the date or dates sampling occurred in YYYY-MM-DD format. If reporting an interval of dates, separate the dates with a slash (/). Please be sure to provide accompanying sampling protocol and effort information in the terms **dwc:samplingProtocol**, **dwc:sampleSizeValue**, **dwc:sampleSizeUnit**, and **dwc:samplingEffort**.
EXAMPLES:
- 2015-08-14
  - For single day sampling.
- 2015-08-14T16:00:00/2015-08-15T16:00:00
  - For sampling over a maximum 24-hour period that spans two dates. This is interpreted as: collections started on August 14, 2015 at 4pm local time and ended August 15, 2015 at 4pm local time. Note that in this case, **dwc:eventTime** does not need to be provided, but can be, as it is included in **dwc:eventDate**.

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:eventDate
https://www.gbif.org/data-quality-requirements-occurrences#dcEventDate

TERM NAME: **eventTime**
DEFINITION: The time or interval during which a sampling event occurred.
RATIONALE: This term is used in conjunction with **dwc:eventDate** to clarify the sampling time interval and is best used when sampling occurs on a single day.
REQUIREMENT: Core.
HOW TO USE: Recommended best practice is to use a time that conforms to ISO 8601-1:2019. Report in local time. Report the total sampling time interval, covering the start and end time of either passive trap deployment or active sampling. The correct format is start time/end time. Use **dwc:samplingEffort** to report the duration of sampling time in hours, person-hours, or decimals of hours or person-hours. Please be sure to provide accompanying sampling protocol and effort information in the terms **dwc:samplingProtocol**, **dwc:sampleSizeValue**, **dwc:sampleSizeUnit**, and **dwc:samplingEffort**.
EXAMPLES:
- 09:00/15:00, 09:30/09:40, 11:50/12:00, 14:15/14:25
  - For a 24-hour interval, please use **dwc:eventDate** instead.
- 08:30/08:30
  - If choosing to report a 24-hour interval.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:eventTime

Term name: **year**
Definition: The four-digit year in which the sampling event occurred, according to the Common Era Calendar.
Rationale: Separating the full **dwc:eventDate** into year, month, and day simplifies data entry and sorting for analysis.
Requirement: Core.
How to use: This is a numeric field. Enter the four-digit year the sampling event occurred.
Examples: 1965, 2013
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:year

Term name: **month**
Definition: The integer month in which the sampling event occurred.
Rationale: Separating the full **dwc:eventDate** into year, month, and day simplifies data entry and sorting for analysis. Bee communities vary widely by month.
Requirement: Core.
How to use: This is a numeric field. Enter the one or two-digit month when the sampling event occurred.
Examples: 1, 10
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:month

Term name: **day**
Definition: The integer day of the month on which the sampling event occurred.
Rationale: Separating the full **dwc:eventDate** into year, month, and day simplifies data entry and sorting for analysis.
Requirement: Core, unless sampling occurred over multiple days.
How to use: This is a numeric field. Enter the one or two-digit day of the month when the sampling event occurred. If sampling occurred over multiple days, *i.e.*, passive traps were used, do not use this term, as it only records one day. The terms **dwc:year** and **dwc:month** should still be used if sampling occurred over multiple days.
Examples: 8, 23
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:day

Term name: **samplingProtocol**
Definition: The names of, references to, or descriptions of the methods or protocols used during a sampling event.
Rationale: Describing the sampling method(s) used is a requirement for some types of species distribution models. It also improves replicability of study methods and provides insights into life-history useful for recollection.
Requirement: Core.
How to use: This term uses two controlled vocabularies; one for active sampling methods and one for passive sampling methods. Additionally, some of the active sampling entries are provided using key:value pairs. Select the sampling method

used from the list of allowable entries.

- Controlled vocabulary lists for **dwc:samplingProtocol**:
  - Active sampling: hand net:lethal, hand net:released, sweep net:lethal, sweep net:released, vial:lethal, vial:released, photograph, visual observation
  - Passive sampling: bowl traps, glycol cups, malaise traps, vane traps

In addition to providing the sampling method(s) following the controlled vocabulary list, cite the source of your protocol following a vertical bar space character (|)

Cariveau, D.P., K.-L.J. Hung, N.M. Williams, D.W. Inouye, C.T. Burns, I.G. Lane, R.E. Irwin, H.K. Levenson, B. Du Clos, & S.H. Woodard. 2025. Standardized protocols for collecting data on bee-flower interactions and the associated floral community. Journal of Melittology 123(5): 104–138. https://doi.org/10.17161/jom.vi123.23861

Levenson, H.K., O. Messinger Carril, N.E. Turley, C. Maffei, G. LeBuhn, T. Griswold, N.M. Williams, K.-L.J. Hung, R.E. Irwin, B. Du Clos, & S.H. Woodard. 2025. Standardized protocol for collecting community-level bee data. *Journal of Melittology* 123(4): 78–103. https://doi.org/10.17161/jom.vi123.22649

López-Uribe, M.M., J.P. Strange, L. Whiteman, B.N. Danforth, S. Jha, M.G. Branstetter, J.B.U. Koch, H.K. Levenson, B. Du Clos, & S.H. Woodard. 2025. Standardized protocol for collecting bee samples to generate molecular data. *Journal of Melittology* 123(7): 163–181. https://doi.org/10.17161/jom.vi123.22596

Otto, C.R.V., L.L. Bailey, B. Du Clos, T. Smith, E. Evans, I. Pearse, S. Killingsworth, S. Jepsen, & S.H. Woodard. 2025. Estimating occupancy of focal bee species. *Journal of Melittology* 123(6): 139–162. https://doi.org/10.17161/jom.vi123.22555

Strange, J.P., M.M. López-Uribe, L. Whiteman, B.N. Danforth, S. Jha, H.K. Levenson, B. Du Clos, J.B.U. Koch, & S.H. Woodard. 2025. Standardized protocol for collecting bee samples for internal parasite and pathogen data. *Journal of Melittology* 123(8): 182–194. https://doi.org/10.17161/jom.vi123.22598

LeBuhn, G., T. Griswold, R. Minckley, S. Droege, T. Roulston, J. Cane, F. Parker, S. Buchmann, V. Tepedino, N. Williams, C. Kremen, & O. Messinger. 2003. A standardized method for monitoring bee populations—the bee inventory (BI) plot. Available from https://www.nativebeemonitoring.org/s/Bee-Plot-2003.pdf

A Collective and Ongoing Collaborative Effort by Those Who Love to Study Bees in North America. (2024). The Very Handy Bee Manual (2.0). Zenodo. https://doi.org/10.5281/zenodo.12812754

Maffei, C., Lent, S., Lane, I., Jones, P., & K. Dillon. 2025. National Protocol Framework for the Inventory and Monitoring of Bees. Version 3.0. Inventory and Monitoring, National Wildlife Refuge System, U.S. Fish and Wildlife Service, Fort Collins, Colorado. https://iris.fws.gov/APPS/ServCat/Reference/Profile/179113

Packer, L., & G. Darla-West. 2021. Bees: How and Why to Sample Them. In: Santos, J.C., Fernandes, G.W. (eds) Measuring Arthropod Biodiversity. Springer, Cham. https://doi.org/10.1007/978-3-030-53226-0_3

Other protocols used can be cited.

Papers published that describe the protocol used to collect the specimen can be cited.

EXAMPLES:

bowl trap | *The Very Handy Bee Manual* (A Collective 2024) https://doi.org/10.5281/zenodo.12812754

hand net:lethal | Maffei et al., (2025) https://iris.fws.gov/APPS/ServCat/Reference/

Profile/179113
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:samplingProtocol

Term name: **sampleSizeValue**
Definition: A numeric value for a measurement of the size (time duration, length, area, or volume) of a sample in a sampling event.
Rationale: Allows for full reporting of sampling effort, supporting analyses of species distribution. This term is used to describe the number of traps used when passive sampling or the number of personnel conducting active sampling in a sampling event.
Requirement: Core.
How to use: Use of this term varies with active or passive sampling methods, but it must always be used with **dwc:sampleSizeUnit**. Provide a number of sampling units specific to passive samples (number of traps collected) or active samples (number of collectors or observers).
Examples: 1, 25
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:sampleSizeValue

Term name: **sampleSizeUnit**
Definition: A unit of measurement of the size of a sample in a sampling event.
Rationale: Allows for full reporting of sampling effort, supporting analyses of species distribution. This term is used to describe the type of equipment used in passive sampling or the role of personnel conducting active sampling during a sampling event.
Requirement: Core.
How to use: This term uses a controlled vocabulary. Use of this term varies with active or passive sampling methods, but it must always be used with **dwc:sampleSizeValue**. Provide a sampling unit specific to passive samples ([type of] traps collected) or active samples (collectors or observers).
- Controlled vocabulary list for **dwc:sampleSizeUnit**:
  - bowl, cup, Malaise, or vane traps collected
  - collectors
  - observers

Examples: Bowl traps collected, cup traps collected, observers
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:sampleSizeUnit

Term name: **samplingEffort**
Definition: The amount of effort expended during a sampling event.
Rationale: Describing the sampling effort expended to obtain a species observation is a requirement for some types of species distribution models.
Requirement: Core.
How to use: This term is used to report both sampling area (in metric units) and duration of sampling time. Report duration of passive sampling time in hours or decimals of hours. Report duration of active sampling time in person-hours (*i.e.*, the amount of work done by one person in one hour) or decimals

of person-hours. If reporting both sampling area and duration of sampling time, separate them with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell. Use **dwc:eventDate** to report the start and end times of sampling.

EXAMPLES:

- 100 m transect, 1 ha, 100m2
  - For reporting amount of area sampled.
- 6 hours, 24 hours
  - When recording duration of passive sampling.
- 0.6 person-hours
  - Total time spent active sampling in person-hours; in this case, 2 people each spent 20 minutes surveying live bees.
- 0.25 person-hours
  - Total time spent active sampling in person-hours; in this case, 1 person spent 15 minutes netting.
- 0.5 ha plot | 0.5 hours
  - Recording both area sampled and duration of passive sampling time.

RELEVANT PROTOCOLS: All.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:samplingEffort

Location

TERM NAME: **country**

DEFINITION: The name of the country or major administrative unit in which the sampling location occurs.

RATIONALE: Providing finer grain location information provides essential context for interpretation and supports data filtering for analysis.

REQUIREMENT: Core.

HOW TO USE: This term uses a controlled vocabulary. See the Getty Thesaurus of Geographic Names (https://www.getty.edu/research/tools/vocabularies/tgn/) for a full controlled vocabulary list.

EXAMPLES: United States, Canada, Mexico

RELEVANT PROTOCOLS: All.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:country

TERM NAME: **stateProvince**

DEFINITION: The name of the next smaller administrative region than the country (state, province, etc.) in which the sampling occurs.

RATIONALE: Providing finer-grain location information provides essential context for interpretation and supports data filtering for analysis.

REQUIREMENT: Core.

HOW TO USE: This term uses a controlled vocabulary. See the Getty Thesaurus of Geographic Names (https://www.getty.edu/research/tools/vocabularies/tgn/) for a full controlled vocabulary list.

EXAMPLES: Kansas, Michigan

RELEVANT PROTOCOLS: All.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:stateProvince

Term name: **decimalLatitude**
Definition: The geographic latitude (in decimal degrees, using the spatial reference system given in **dwc:geodeticDatum**) of the geographic center of a sampling location.
Rationale: A precise location of a wild bee observation provides essential context for interpretation. Providing a precise location supports species status assessments, distribution models, and other analyses relying on occurrence location.
Requirement: Core.
How to use: Positive values are north of the Equator, negative values are south of it. Legal values lie between–90 and 90, inclusive. If coordinates are obtained from a phone or non-professional grade GPS then the accuracy is likely, at best, around 5 m, which would be best reflected in reporting 5 decimal places. If coordinates are derived from geographic computer software (*i.e.*, Google Maps or GIS programs), they may have as many as six decimal places.
Examples:–41.09837
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:decimalLatitude

Term name: **decimalLongitude**
Definiion: The geographic longitude (in decimal degrees, using the spatial reference system given in **dwc:geodeticDatum**) of the geographic center of a sampling location.
Rationale: A precise location of a wild bee observation provides essential context for interpretation. Providing a precise location supports species status assessments, distribution models, and other analyses relying on occurrence location.
Requirement: Core.
How to use: Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between–180 and 180, inclusive. If coordinates are obtained from a phone or non-professional grade GPS then the accuracy is likely, at best, around 5 m, which would be best reflected in reporting 5 decimal places. If coordinates are derived from geographic computer software (*i.e.*, Google Maps or GIS programs), they may have as many as six decimal places.
Examples:–121.17616
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:decimalLongitude

Term name: **coordinateUncertaintyInMeters**
Definition: The horizontal distance (in meters) from the given **dwc:decimalLatitude** and **dwc:decimalLongitude** describing the smallest circle containing the sampling location.
Rationale: A precise location of a wild bee observation supports species status assessments, distribution models, and other analyses relying on occurrence location. Sometimes the location of an observation must be masked for species or other security; the extent of masking can be provided here.
Requirement: Core.
How to use: Report this value if coordinates are obtained from a GPS device, including a phone or non-professional grade GPS. The coordinate uncertainty for these devices is likely around 5 meters. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term. This value, in most cases for

current and future occurrences, should not be greater than 10. Exceptions include if the occurrence location is masked for security reasons or if the most accurate coordinates available are less precise, as is the case for much historically-collected occurrence data.

EXAMPLES:

- 5
  - When precise GPS-derived coordinates with five decimal places are provided (see **dwc:decimalLatitude** and **dwc:decimalLongitude**).
- 0.1
  - When coordinates are derived from geographic computer software (i.e., Google Maps or GIS programs) with six decimal places.
- 100, 1000
  - When occurrence location is masked by 100 meters, if providing GPS-derived coordinates with three decimal places, or 1 km, if providing GPS-derived coordinates with two decimal places. Report any location masking, if applicable, in **dwc:informationWithheld**.

RELEVANT PROTOCOLS: All.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:coordinateUncertaintyInMeters https://www.gbif.org/data-quality-requirements-occurrences#dcUncertainty

Identification

TERM NAME: **identifiedBy**

DEFINITION: A list of names of people, groups, or organizations who identified the specimen associated with an occurrence.

RATIONALE: Providing the names of the identifying personnel creates a potential point of contact for more taxonomic information regarding an occurrence.

REQUIREMENT: Core.

HOW TO USE: Use full names (First [Middle Initial, if provided] Last) whenever possible. Separate the values in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell. This term can be used in conjunction with **dwc:identifiedByID** to connect the occurrence to the identifying personnel through platforms such as Bionomia.

EXAMPLES:

- For one identifier:
  - Erika M. Tucker
- For more than one identifier:
  - Theodore Pappenfuss | Robert Macey

RELEVANT PROTOCOLS: All.

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:identifiedBy

Taxon

TERM NAME: **scientificName**

DEFINITION: The full scientific name, with authorship and date information if known.

RATIONALE: This term is required for publication to GBIF. Further, accurate

identification to the lowest taxonomic rank possible creates the most useful data from a set of occurrences.
Requirement: Core.
How to use: This should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the **dwc:identificationQualifier** term. Authorship of the scientific name, including the date, can also be provided separately in **dwc:scientificNameAuthorship**.
Examples: Hymenoptera, Halictidae, *Sphecodes* sp., *Osmia atriventris* (Cresson, 1864).
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:scientificName
https://www.gbif.org/data-quality-requirements-occurrences#dcSciName

Term name: **genus**
Definition: The full scientific name of the genus in which the occurrence is classified.
Rationale: Separating the full **dwc:scientificName** into genus and specific epithet simplifies data entry and sorting for analysis.
Requirement: Core.
How to use: Provide the genus of the bee with the first letter capitalized.
Examples: *Andrena*, *Coelioxys*
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:genus

Term name: **specificEpithet**
Definition: The name of the first or species epithet of the **dwc:scientificName**.
Rationale: Separating the full **dwc:scientificName** into genus and specific epithet simplifies data entry and sorting for analysis.
Requirement: Core.
How to use: Provide the specific epithet of the bee with the first letter lowercase. If not available, specify sp.
Examples: *mirabilis*, *carlini*, sp.
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:specificEpithet

Recommended Terms

Record-level

Term name: **license**
Definition: A document indicating the preferred terms of data use by the data provider.
Rationale: Assigning a license to a dataset informs users what can be done with that dataset and the appropriate crediting procedure. Some data providers do not want their data used for commercial purposes and may specify that with their license choice.
Requirement: Recommended.
How to use: Choose a Creative Commons license that aligns with the desired potential use cases of the dataset and crediting criteria of the provider. Provide a

link to the full license terms. Licenses to use include: public domain, CC-BY, CC-BY-SA, CC-BY-NC, or CC-BY-NC-SA. You may provide the name of the license holder with the optional term **dwc:rightsHolder**.
EXAMPLES: https://creativecommons.org/publicdomain/zero/1.0/, https://creativecommons.org/licenses/by/4.0/, https://creativecommons.org/licenses/by-sa/4.0/
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dcterms:license

TERM NAME: **institutionID**
DEFINITION: An identifier for the institution having custody of the object(s) or information referred to in the record.
RATIONALE: Provides a location where occurrence information came from; may be used to find a point of contact for more information about an occurrence.
REQUIREMENT: Recommended.
HOW TO USE: If available, use identifiers from the Global Registry of Scientific Collections (https://scientific-collections.gbif.org/).
EXAMPLES: http://grscicoll.org/institution/american-museum-natural-history
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:institutionID

TERM NAME: **collectionID**
DEFINITION: An identifier for the collection or dataset from which the record was derived.
RATIONALE: Provides more precise information regarding the location where occurrence information came from; may be used to find a point of contact for more information about an occurrence.
REQUIREMENT: Recommended.
HOW TO USE: If available, use identifiers from the Global Registry of Scientific Collections (https://scientific-collections.gbif.org/).
EXAMPLES: http://grbio.org/cool/je3k-bvrg, http://grscicoll.org/institutional-collection/entomology
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:collectionID

TERM NAME: **collectionCode**
DEFINITION: The name, acronym, code, or initialism identifying the collection or dataset from which the record was derived.
RATIONALE: Provides more precise information regarding the location where occurrence information came from; may be used to find a point of contact for more information about an occurrence.
REQUIREMENT: Recommended.
HOW TO USE: Provide a name or acronym for the collection within the institution or dataset within a lab that houses the specimen or coordinated the observation. If the collection is registered in the Global Registry of Scientific Collections (https://scientific-collections.gbif.org/), please use the code listed there.
EXAMPLES: BBSL, CUIC
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:collectionCode

Occurrence

Term name: **recordedByID**
Definition: A list of the globally unique identifier(s) for the person, people, groups, or organizations responsible for recording the original occurrence.
Rationale: A unique identifier for a data collector, such as an ORCID iD, can be used to aggregate contributions to natural history collections on platforms such as Bionomia.
Requirement: Recommended.
How to use: Recommended best practice is to provide a single identifier that disambiguates the details of the identifying agent. If a list is used, separate the values in the list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
Examples:

- For one collector:
  - https://orcid.org/0000-0002-1825-0097
- For more than one collector:
  - https://orcid.org/0000-0002-1825-0097
  - https://orcid.org/0000-0002-1825-0098

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:recordedByID

Term name: **sex**
Definition: The sex of the biological individual(s) represented in the occurrence.
Rationale: Providing demographic information can improve and inform general knowledge on bee life history habits and species distribution models, particularly for at-risk species.
Requirement: Recommended.
How to use: This term uses a controlled vocabulary. Provide the appropriate category following the controlled vocabulary list.

- Controlled vocabulary list for **dwc:sex**:
  - female
  - male
  - gynandromorph

Examples: Female, male, gynandromorph
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:sex

Term name: **associatedTaxa**
Definition: A list of identifiers or names of taxa and the associations of this occurrence to each of them.
Rationale: Providing the plant a bee was observed on can inform plant-pollinator networks, species conservation plans, and habitat management.
Requirement: Recommended for any active sampling conducted on bees visiting blooming flowers.
How to use: This term uses a controlled vocabulary. Using a key:value pair,

provide the appropriate relationship from the controlled vocabulary list, using a combination of an action and a part of the plant where the bee was found. Provide at least the genus name of the plant, though the full plant scientific name (Genus species) is preferred. We advise reporting the source for the taxonomic name with the authority key:value pair. If using both the plant species and the authority key:value pairs, separate them with a comma. Enclose the entire term with curly brackets ({}, see Examples) :

- Controlled vocabulary list for **dwc:associatedTaxa**:
    - Action vocabulary: "caught on", "observed on", "visits"
    - Plant part vocabulary: "flowers of", "leaves of", "stem of"

EXAMPLES:

- {"visits flowers of":"*Rubus*"}
    - For reporting plant interaction only and identifying plant to genus level.
- {"observed on flowers of":"*Solidago canadensis*", "authority":"example authority"}
    - For reporting plant interaction, full scientific name of the plant, and the source of the taxonomic identification of the plant.
    - The authority should be a full citation of the source used to identify the plant. Examples include:
        - Flora Novae Angliae. 2011. Yale University Press https://plants.usda.gov/home/plantProfile?symbol=SOCA6
        - The Jepson Manual: Vascular Plants of California. Second Edition, 2012. University of California Press
        - For further details on reporting taxonomic authority for species identifications, see **dwc:namePublishedIn**.

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:associatedTaxa

MaterialEntity

TERM NAME: **disposition**
DEFINITION: The current state of a specimen with respect to the collection identified in **dwc:collectionCode** or **dwc:collectionID**.
RATIONALE: This term can indicate the availability of a specimen for further study.
REQUIREMENT: Recommended.
HOW TO USE: This term uses a controlled vocabulary. Specify the appropriate **dwc:disposition** from the controlled vocabulary list. If using different terms, please define them in **dwc:materialEntityRemarks**.

- Controlled vocabulary list for **dwc:disposition**:
    - inCollection: specimen is preserved in a collection.
    - missing: specimen is missing from the collection.
    - onLoan: specimen is on loan to another institution, organization, or individual.
    - destroyed: specimen has been destroyed.

EXAMPLES:

- inCollection
- destroyed
- Awaiting processing (with details provided in **dwc:materialEntityRemarks**)

- Missing head (with details provided in **dwc:materialEntityRemarks**)

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:disposition

Event

Term name: **eventID**
Definition: An identifier for the set of information associated with a sampling event. May be a global unique identifier or an identifier specific to the dataset.
Rationale: When digitizing data and connecting sampling events to occurrence records, a **dwc:eventID** can be linked to a separate event table, eliminating redundancy throughout an occurrence table.
Requirement: Recommended.
How to use: Create a unique identifier for a sampling event by combining site, date, and location information associated with that event. See the section "Generating unique identifiers."
Examples:

- 0320190603
    - If you sampled site 3 on June 3, 2019, with 03 coming from site 3 and the remainder of the ID coming from the date formatted for **dwc:eventDate**.

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:eventID

Term name: **eventRemarks**
Definition: Comments or notes about the sampling event.
Rationale: Provides additional context regarding the sampling event. This can include, but is not limited to, notes on present or historical field conditions, collector experience, or additional information on passive trap deployment.
Requirement: Recommended.
How to use: Provide any additional information about the sampling event. Separate entries with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell. Field condition information can include habitat change such as mowing, fire, harvesting, etc., or pesticide application history. To report collector experience, self-identify as: Novice, Advanced Beginner, Competence, Proficient, or Expert. Additional information on passive trap deployment can include the trap colors used, trap size, height at which traps were deployed, or trap liquid type. Additional information on transects used for active sampling can indicate whether the transect was linear or meandering.
Examples:

- Site planted with seeds on 2022-04-29 | collectorExperience: Advanced Beginner | meandering transect
    - Describes an active sampling event at a site where seeds were planted on April 29, 2022. The collector, who has some experience, walked a meandering transect as they sampled.
- Blue, yellow, and white traps used | 3.25 oz cups | Dawn dish soap | site recently mowed
    - Describes a passive sampling event where all three commonly used trap colors were deployed using soapy water in Solo souffle cups at a site that was recently mowed.

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:eventRemarks

Location

TERM NAME: **locationID**
DEFINITION: An identifier for the set information about a sampling location. May be a global unique identifier or an identifier specific to the dataset.
RATIONALE: When digitizing data and connecting sampling locations to occurrence records, a **dwc:locationID** can be linked to a separate location table, eliminating redundancy throughout an occurrence table.
REQUIREMENT: Recommended.
HOW TO USE: Create a unique identifier for a sampling location with information associated with that location, including a site number, coordinates, or place name. See the section "Generating unique identifiers."
EXAMPLES:
- Site09, site15, site3A
    - Site numbers can also be provided in **dwc:fieldNumber**.

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:locationID

TERM NAME: **geodeticDatum**
DEFINITION: The ellipsoid, geodetic datum, or spatial reference system upon which the geographic coordinates given in **dwc:decimalLatitude** and **dwc:decimalLongitude** are based.
RATIONALE: Providing a datum gives context to provided coordinates. GBIF assumes the datum is WGS84. If you don't know what coordinate system you are using it is most likely WGS84, which is the default for publicly available maps like Google Maps and is what is used for Global Positioning Systems (GPS). If coordinates were provided from a different datum, that should be specified. For instance, NAD27 was once a common datum, but coordinates in NAD27 can be up to 200 meters away from those in WGS84.
REQUIREMENT: Recommended.
HOW TO USE: The best practice is to use the EPSG code for the datum or spatial reference system. The EPSG code for WGS84 is EPSG:4326 (https://epsg.io/4326). If an EPSG code is not available, the name of the datum or spatial reference system may be provided.
EXAMPLES: EPSG:4326, EPSG:4269, WGS84, NAD27, NAD83
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:geodeticDatum
https://www.gbif.org/data-quality-requirements-occurrences#dcGeodeticDatum

Identification

TERM NAME: **verbatimIdentification**
DEFINITION: A string representing the taxonomic identification as it appeared in the original record. This term is meant to be used in addition to **dwc:scientificName**, not instead of it.
RATIONALE: This term can be used to translate original label text on preserved

specimens, to store the original identification if the **dwc:scientificName** changes, or in any other case where the original identification is different from the current identification.
Requirement: Recommended.
How to use: Translate label text verbatim when digitizing.
Examples: *Osmia besseyae* Cockerell TYPE (https://library.big-bee.net/portal/collections/individual/index.php?occid=1667260)
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:verbatimIdentification

Term name: **identificationQualifier**
Definition: A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the identification.
Rationale: This term can provide more context about a taxonomy of an occurrence.
Requirement: Recommended, if applicable. Must be used with **dwc:taxonRank**.
How to use: This term uses a controlled vocabulary. Provide the appropriate qualifier, if applicable.

- Controlled vocabulary list for **dwc:identificationQualifier** (definitions taken from Sigovini *et al.*, 2016):
  - Affinis (aff.): affinity with a known species; has affinity with
  - Confer (cf.): to compare or be compared with
  - Species incerta (?, sp. inc, or inc.): uncertain species
  - Species proxima (prox.): the nearest species
  - Species near (nr.): near but not identical to a species
  - Stetit (stet.): further identification has not been attempted

Examples:

- cf. *vincta*
  - (for *Nomada* cf. *vincta* with accompanying values *Nomada* in **dwc:genus**, vincta in **dwc:specificEpithet**, and cf. species in **dwc:taxonRank**).
- aff. *modestus*
  - (for *Hylaeus* aff. *modestus* with accompanying values *Hylaeus* in **dwc:genus**, *modestus* in **dwc:specificEpithet**, and aff. species in **dwc:taxonRank**).

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:identificationQualifier

Term name: **identifiedByID**
Definition: A list of the globally unique identifier(s) for the person, people, groups, or organizations responsible for identifying the specimen associated with an occurrence.
Rationale: A unique identifier for bee identifying personnel, such as an ORCID iD, can be used to aggregate contributions to natural history collections on platforms such as Bionomia.
Requirement: Recommended.
How to use: Recommended best practice is to provide a single identifier that disambiguates the details of the identifying agent. If a list is used, separate the values in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.

EXAMPLES:
- For one collector:
  - https://orcid.org/0000-0002-1825-0097
- For more than one collector:
  - https://orcid.org/0000-0002-1825-0097 | https://orcid.org/0000-0002-1825-0098

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:identifiedByID

TERM NAME: **dateIdentified**
DEFINITION: The date a specimen associated with an occurrence was identified.
RATIONALE: Provides context for the taxonomic identification.
REQUIREMENT: Recommended.
HOW TO USE: Provide the date the taxonomic identification was determined, using a date that conforms to ISO 8601-1:2019.
EXAMPLES: 2020-08-09, 1987-04, 1963
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:dateIdentified

TERM NAME: **identificationReferences**
DEFINITION: A list of references used to identify a specimen.
RATIONALE: Provides context and improves reproducibility for the taxonomic identification.
REQUIREMENT: Recommended.
HOW TO USE: Provide citations for references used to identify specimens, separating the values in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
EXAMPLE: Portman, Z.M., M. Arduser, I.G. Lane, & D.P. Cariveau. 2022. A review of the *Augochloropsis* (Hymenoptera, Halictidae) and keys to the shiny green Halictinae of the midwestern United States. *ZooKeys* 1130: 103–152. https://doi.org/10.3897/zookeys.1130.86413 | https://www.discoverlife.org/mp/20q?guide=Bee_genera
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:identificationReferences

Taxon

TERM NAME: **taxonID**
DEFINITION: An identifier for the information describing the taxonomy of an occurrence. May be a global unique identifier or an identifier specific to the dataset.
RATIONALE: When digitizing data and connecting species identifications to occurrence records in a relational database or an Excel workbook, a **dwc:taxonID** can be linked to a separate taxon table, streamlining an occurrence table.
REQUIREMENT: Recommended.
HOW TO USE: See the section "Generating unique identifiers." There are multiple sources for **dwc:taxonID**s for bees; these include species pages on GBIF or DiscoverLife and ITIS taxonomic serial numbers. Any or all of these identifiers

can be provided in **dwc:taxonID**; separate multiple identifiers in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
Example: https://www.gbif.org/species/5042859 | Taxonomic Serial No.: 757559 | https://www.discoverlife.org/20/q?search=Agapostemon+texanus
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:taxonID

Term name: **nameAccordingTo**
Definition: The reference to the source in which the specific taxon concept circumscription is defined or implied, traditionally signified by the Latin "sensu" or "sec." (from secundum, meaning "according to").
Rationale: Providing taxonomic reference information allows occurrence identifications to be verified or examined in further study.
Requirement: Recommended.
How to use: For taxa that result from identifications, a reference to the keys, monographs, experts and other sources should be given.
Example: Ascher, J.S., & J. Pickering. 2024. Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). http://www.discoverlife.org/mp/20q?guide=Apoidea_species
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:nameAccordingTo

Term name: **family**
Definition: The full scientific name of the family in which the occurrence is classified.
Rationale: Provides important taxonomic context and supports data filtering for analysis.
Requirement: Recommended.
How to use: Provide the family name.
Examples: Halictidae, Colletidae
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:family

Term name: **infraspecificEpithet**
Definition: The name of the lowest or terminal infraspecific epithet of the **dwc:scientificName**, excluding any rank designation.
Rationale: Provides important taxonomic context. Accurate identification to the lowest taxonomic rank possible creates the most useful data from a set of occurrences.
Requirement: Recommended.
How to use: Provide the subspecies name.
Examples:

- *virginica*
  - for *Xylocopoides virginica virginica*
- *propinqua*
  - for *Osmia lignaria propinqua*

RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:infraspecificEpithet

TERM NAME: **scientificNameAuthorship**
DEFINITION: The authorship information for the **dwc:scientificName**
RATIONALE: Provides important taxonomic context.
REQUIREMENT: Recommended.
HOW TO USE: Provide the authorship and date of the scientific name.
EXAMPLE: Cockerell, 1906
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:scientificNameAuthorship

Optional Terms

Record-level

TERM NAME: **rightsHolder**
DEFINITION: A person or organization owning or managing rights over the resource.
RATIONALE: Provides contact information regarding use of shared data and license choice.
REQUIREMENT: Optional.
HOW TO USE: Provide the name(s) of the person, organization, or institution responsible for the rights of a shared dataset. Separate multiple values with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
EXAMPLE: The Regents of the University of California
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dcterms:rightsHolder

TERM NAME: **dynamicProperties**
DEFINITION: A list of additional measurements, facts, characteristics, or assertions about the record. This term is meant to provide a mechanism for structured content.
RATIONALE: Reporting site conditions during sampling events can inform occupancy models. Darwin Core does not currently have terms for these conditions; therefore, *The Wild Bee Data Standard* uses **dwc:dynamicProperties** to report them.
REQUIREMENT: Optional.
HOW TO USE: This term uses a controlled vocabulary. Report any or all of the following site conditions using key:value pairs. Note that numeric values do not need quotation marks. If providing multiple key:value pairs, separate them with commas. Enclose the entire term in curly brackets ({}, see Example).

- Controlled vocabulary list for **dwc:dynamicProperties**:
  - airTemperature: report degrees Celsius
  - relHumidity: report relative humidity percent value
  - windSpeed: report wind speed and unit measure
  - cloudCover: report one of the following: clear, partly cloudy, mostly

cloudy, light overcast, dark overcast
    - precip: report one of the following: none, light rain
        - Note light rain is only allowed for long term passive traps, such as cup or vane traps. Nearly all wild bee sampling should be conducted with no precipitation.
    - AQI: air quality index as reported for the sampling area

Example: {"relHumidity":28, "airTemperature":22, "windSpeed":"3 kph", "cloudCover":"light overcast", "precip":"none", "AQI":34}
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:dynamicProperties

Occurrence

Term name: **lifeStage**
Definition: The age class or life stage of the organism(s) at the time the occurrence was recorded.
Rationale: Providing demographic information can improve and inform general knowledge on bee life history habits and species distribution models, particularly for at-risk species.
Requirement: Optional, but required if the specimen is not an adult.
How to use: This term uses a controlled vocabulary. Provide the appropriate category following the controlled vocabulary list.

- Controlled vocabulary list for **dwc:lifeStage**:
    - egg
    - larva
    - pupa
    - adult

Examples: Larva, adult
Relevant protocols: All, particularly the occupancy of focal species protocol (Otto *et al.*, 2025).
For more detail: https://dwc.tdwg.org/terms/#dwc:lifeStage

Term name: **caste**
Definition: Categorization of individuals for eusocial species.
Rationale: Providing demographic information can improve and inform general knowledge on bee life history habits and species distribution models, particularly for at-risk species.
Requirement: Optional.
How to use: This term uses a controlled vocabulary. Provide the appropriate category following the controlled vocabulary list.

- Controlled vocabulary list for **dwc:caste**:
    - queen
    - worker
    - drone

Examples: Queen, worker
Relevant protocols: All, particularly the occupancy of focal species protocol (Otto *et al.*, 2025).

FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:caste

TERM NAME: **behavior**
DEFINITION: The behavior shown by the subject at the time the occurrence was recorded.
RATIONALE: Providing demographic information can improve and inform general knowledge on bee life history habits and species distribution models, particularly for at-risk species.
REQUIREMENT: Optional.
HOW TO USE: This term uses a controlled vocabulary. Provide the appropriate category following the controlled vocabulary list.
- Controlled vocabulary list for **dwc:behavior**:
  - foraging
  - collecting pollen
  - collecting nectar
  - nesting

EXAMPLES: Collecting pollen, nesting
RELEVANT PROTOCOLS: All, particularly the occupancy of focal species protocol (Otto *et al.*, 2025).
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:behavior

TERM NAME: **vitality**
DEFINITION: An indication of whether an organism was alive or dead at the time of collection or observation.
RATIONALE: Providing demographic information can improve and inform general knowledge on bee life history habits and species distribution models, particularly for at-risk species.
REQUIREMENT: Optional.
HOW TO USE: This term uses a controlled vocabulary. Provide the appropriate category following the controlled vocabulary list.
- Controlled vocabulary list for **dwc:vitality**:
  - alive
  - dead
  - moribund

EXAMPLES: Alive, dead
RELEVANT PROTOCOLS: All, particularly the occupancy of focal species protocol (Otto *et al.*, 2025).
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:vitality

TERM NAME: **associatedMedia**
DEFINITION: A list of identifiers of media associated with the occurrence.
RATIONALE: Provides links to accompanying photos related to the occurrence, including but not limited to site photos, plant photos, trap photos, specimen photos. May instead provide a host institution housing non-public image data or other media.
REQUIREMENT: Optional.
HOW TO USE: Media shared publicly via **dwc:associatedMedia** must have an associated identifier, whether that is a website, a publication, or a UUID. Please

provide any relevant identifiers, separated by a vertical bar if necessary. For non-publicly held media, please provide the name of the institution that manages the media data.

Examples: https://www.flickr.com/photos/usgsbiml/52264266775/, University of Minnesota

Relevant protocols: All.

For more detail: https://dwc.tdwg.org/terms/#dwc:associatedMedia

Term name: **associatedOccurrences**

Definition: A list of identifiers of other occurrence records and their associations to this occurrence

Rationale: Links to other bee or plant occurrences that were observed, detected, or collected around a bee occurrence. This differs from **dwc:associatedTaxa** in that a plant listed here can be another plant in the vicinity of the plant the occurrence was directly interacting with. However, this term can also be used to link to the occurrence of the plant described in **dwc:associatedTaxa**. Use of this term may provide more context for focal species analyses.

Requirement: Optional.

How to use: Use a key:value pair, with the key describing the relationship with the associated occurrence and the value being the unique ID for that occurrence. The unique ID will ideally be a URL link to the occurrence provided on a data portal. We do not provide a controlled vocabulary for the possible relationships; be concise but descriptive when generating these keys. The unique ID can describe for another bee or plant occurrence that a bee occurrence was observed, detected, or collected on. It could also describe the occurrence itself with the use of the "same as" or "same occurrence as" key. If providing multiple key:value pairs, separate them with a comma. Enclose the entire term in curly brackets ({}, see Examples).

Examples:

- {"parasitized by":https://www.gbif.org/occurrence/2851169659}
    - Describes an associated occurrence of a parasitic bee.
- {"observed near": https://www.inaturalist.org/observations/216532139}
    - Describes an associated occurrence of the plant a bee was observed on.
- {"same as": https://www.inaturalist.org/observations/220375291} or {"same occurrence as":https://www.inaturalist.org/observations/220375291}
    - Describes the bee occurrence as posted on iNaturalist. This example can also be used for plant occurrences described in **dwc:associatedTaxa**.

Relevant protocols: All.

For more detail: https://dwc.tdwg.org/terms/#dwc:associatedOccurrences

Term name: **occurrenceRemarks**

Definition: Comments or notes about the occurrence.

Rationale: This term provides additional context for the corresponding occurrence. Here, the term is used to describe how a bee was captured.

Requirement: Optional.

How to use: Provide a brief description of how and where the bee was captured or observed.

Examples: netted in air, netted on the ground

RELEVANT PROTOCOLS: All, particularly the communities protocol (Levenson *et al*., 2025a).
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:occurrenceRemarks

MaterialEntity

TERM NAME: **preparations**
DEFINITION: A list of preparations and preservation methods for material associated with an occurrence.
RATIONALE: Describes procedures taken to preserve material from a physical specimen for molecular or tissue analyses. Can also be used to reference whole specimen preparation.
REQUIREMENT: Core for molecular (López-Uribe *et al*., 2025) and parasite and pathogen (Strange *et al*., 2025) protocols, optional otherwise.
HOW TO USE: Indicate what, if any, material was extracted from the specimen and what that material will be used for. For whole specimens, indicate their final preservation method.
EXAMPLES: DNA extract,–80 freezer, pinned, in alcohol
RELEVANT PROTOCOLS: All, particularly the molecular (López-Uribe *et al*., 2025) and parasite and pathogen (Strange *et al*., 2025) protocols.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:preparations

TERM NAME: **associatedSequences**
DEFINITION: A list of identifiers of genetic sequence information associated with an occurrence.
RATIONALE: Links to genetic sequence data to verify taxonomic identification for a specimen.
REQUIREMENT: Optional.
HOW TO USE: Provide one or multiple means to locate genetic sequence information. Identifiers can include publications, globally unique identifiers, and URLs. Separate identifiers in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
EXAMPLES: https://www.boldsystems.org/index.php/Public_RecordView?processid=BEECC462-08, https://www.boldsystems.org/index.php/Public_RecordView?processid=BUSA207-05
RELEVANT PROTOCOLS: All, particularly the molecular protocol (López-Uribe *et al*., 2025).
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:associatedSequences

TERM NAME: **materialEntityRemarks**
DEFINITION: Comments or notes about the MaterialEntity.
RATIONALE: This term provides context for other terms used in the MaterialEntity category. In *The Wild Bee Data Standard*, these terms are **dwc:disposition**, **dwc:preparations**, and **dwc:associatedSequences**.
REQUIREMENT: Core for molecular (López-Uribe *et al*., 2025) and parasite and

pathogen (Strange *et al.*, 2025) protocols, optional otherwise.
How to use: Use this term to provide information related to **dwc:disposition** that is not included in the controlled vocabulary list. This term is also used in the Bee Monitoring RCN protocols to provide more context for any material taken for molecular or tissue analyses for which the final storage conditions are described in **dwc:preparations**.
Examples:

- in lab
    - Example related to **dwc:disposition** that indicates the specimen is in the laboratory awaiting full processing.
- sterilized with 10% bleach | stored on ice between collection and final storage | 0.8 hour between collection and final storage | stored in Sample Lab, Sample Location
    - Example specific to Bee Monitoring RCN protocols for molecular and tissue data analysis.
- Head is missing from specimen
    - Example related to **dwc:disposition** that provides a note about specimen condition.
- Leg removed for molecular analysis
    - Example related to **dwc:disposition** that provides a note about specimen condition.

Relevant protocols: All, particularly the molecular (López-Uribe *et al.*, 2025), and parasite and pathogen (Strange *et al.*, 2025) protocols.
For more detail: https://dwc.tdwg.org/terms/#dwc:materialEntityRemarks

Event

Term name: **fieldNumber**
Definition: An identifier for a sampling event in the field that links field-based or collected information to an occurrence. Examples include a site number or a vial number associated with a bee or group of bees collected at a site. Note that it is possible for multiple vials to be collected at a site during a sampling event. Using this term to indicate site number, vial number, or some other field-based identifier is allowable.
Rationale: Clarifies sampling protocol.
Requirement: Optional.
How to use: Provide any site-identifying information here.
Examples:

- Site numbers: Site 1, 5A-2015, Farm Site 6
- Vial numbers: 2, AM-Site4-08

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:fieldNumber

Term name: **habitat**
Definition: A category or description of the habitat in which the sampling event occurred.
Rationale: Habitat type has ecological implications on bee community assemblage.
Requirement: Optional.

How to use: Provide any description of the habitat type where the specimen was observed or collected. For a common reference, provide a habitat type and number from either the NRCS Ecological Site Descriptions (ESD) or the EPA Level 3 Ecoregions, specifying which source was used using a key:value pair. If using a key:value pair, enclose the entire term in curly brackets ({}, see Examples).

- More on NRCS ESDs: https://www.nrcs.usda.gov/getting-assistance/technical-assistance/ecological-sciences/ecological-site-descriptions
- More on EPA Ecoregions: https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states

Examples: oak savanna, steppe, {"EPA Level 3 Ecoregion":"Ozark Highlands (39)"}, {"NRCS ESD":"Snake River Plains (011X)"}
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:habitat

Location

Term name: **county**
Definition: The full, unabbreviated name of the next smaller administrative region than stateProvince (county, department, etc.) in which the sampling location occurs.
Rationale: Providing finer grain location information supports data filtering for analysis.
Requirement: Optional.
How to use: This term uses a controlled vocabulary. See the Getty Thesaurus of Geographic Names (https://www.getty.edu/research/tools/vocabularies/tgn/) for a full controlled vocabulary list.
Examples: Ness, Menominee
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:county

Term name: **locality**
Definition: The specific description of the place.
Rationale: Providing finer grain location information supports data filtering for analysis.
Requirement: Optional.
How to use: This could be any useful collection location name such as: town, park, study site code name, street address, or an amalgamation of other location names in your dataset.
Examples: Grand Staircase-Escalante National Monument, 500 m from Hwy 1 at Old Orchard Rd
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:locality

Term name: **verbatimElevation**
Definition: The original description of the elevation (altitude, usually above sea level) of the sampling location.
Rationale: Elevation has ecological implications on bee community assemblage.

Requirement: Optional.
How to use: Provide an estimate of elevation above sea level in meters. The value can be gathered in the field with GPS or via spatial software tools.
Example: 150 m
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:verbatimElevation

Term name: **coordinatePrecision**
Definition: A decimal representation of the precision of the coordinates given in the **dwc:decimalLatitude** and **dwc:decimalLongitude**.
Rationale: Providing precision is another way of stating uncertainty in the coordinates in conjunction with the core term **dwc:coordinateUncertaintyInMeters**. It can also clarify rounding errors that may occur in computer software when the end of a coordinate value is zero.
Requirement: Optional.
How to use: Provide the appropriate number of decimals in the coordinates of the occurrence record.
Examples:

- 0.00001
    - precise to 5 m; best practice for *The Wild Bee Data Standard*. Associated with GPS-derived coordinates with five decimal places.
- 0.000001
    - precise to 0.11 m; associated with coordinates with six decimal places derived from geographic computer software (*i.e.*, Google Maps, GIS programs).
- 0.01
    - precise to 1 km. The value for **dwc:coordinateUncertaintyInMeters** will be 1000, and the number of decimal places in **dwc:decimalLatitude** and **dwc:decimalLongitude** will be 1. Be sure to indicate location masking in **dwc:informationWithheld**.

Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:coordinatePrecision

Term name: **georeferencedBy**
Definition: A list of names of people, groups, or organizations who determined the georeference (spatial representation) for the sampling location.
Rationale: Providing the names of the georeferencing personnel creates a potential point of contact for more information regarding an occurrence or its collecting event
Requirement: Optional.
How to use: Use full names (First [Middle Initial, if provided] Last) whenever possible. Separate the values in a list with space vertical bar space (|); this separator is used throughout the Darwin Core standard to concatenate multiple entries in a cell.
Example: Amelia Earhart | Harry Manning
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:georeferencedBy

Term name: **georeferenceRemarks**
Definition: Notes or comments about the spatial description determination.

RATIONALE: Some organizations require detailed information regarding the georeference process, including how many satellites triangulated the coordinates or the type of device used to determine the coordinates.
REQUIREMENT: Optional
HOW TO USE: Provide any relevant information to satisfy organization requirements, including the number of satellites or the device type used to record coordinates. To clarify to users outside your organization what information is being provided, describe the information using a key:value pair. Note that numerical values do not need quotation marks. If using multiple key:value pairs, separate them with a comma. Enclose the entire term in curly brackets ({}, see Example).
EXAMPLE: {"number of satellites":5, "device type used to record coordinates": "Garmin eTrex 22x"}
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:georeferenceRemarks

Identification

TERM NAME: **typeStatus**
DEFINITION: A list of nomenclatural types (type status, typified scientific name, publication) applied to the subject.
RATIONALE: Provides important taxonomic context by allowing for identification verification.
REQUIREMENT: Optional if not a type specimen, but required if the occurrence record refers to a type specimen.
HOW TO USE: Provide any type status information, if applicable.
EXAMPLE: holotype of *Halictus confusus*
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:typeStatus

TERM NAME: **identificationRemarks**
DEFINITION: Any comments or notes regarding the identification of a specimen.
RATIONALE: Provides context and improves reproducibility for taxonomic identification.
REQUIREMENT: Optional.
HOW TO USE: Provide any ancillary information that may be helpful to others attempting to verify, modify, or study the identification of a specimen.
EXAMPLE: Made determination based on malar space
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:identificationRemarks

Taxon

TERM NAME: **namePublishedIn**
DEFINITION: A reference for the publication in which the **dwc:scientificName** was originally established.
RATIONALE: Provides important taxonomic context and improves reproducibility for

taxonomic identification by allowing for reference verification by data users.
Requirement: Optional.
How to use: Provide a full citation for a taxonomic reference.
Example: Sandhouse, G.A. 1937. The bees of the genera *Augochlora*, *Augochloropsis*, and *Augochlorella* (Hymenoptera; Apoidea) occurring in the United States. *Journal of the Washington Academy of Sciences* 27: 65–79.
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:namePublishedIn

Term name: **tribe**
Definition: The full scientific name of the tribe in which the occurrence is classified.
Rationale: Provides important taxonomic context. Accurate identification to the lowest taxonomic rank possible creates the most useful data from a set of occurrences.
Requirement: Optional.
How to use: Provide the tribe name.
Examples: Colletini, Perditini, Nomiini
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:tribe

Term name: **subgenus**
Definition: The full scientific name of the subgenus in which the occurrence is classified. Values should include the genus to avoid homonym confusion.
Rationale: Provides important taxonomic context. Accurate identification to the lowest taxonomic rank possible creates the most useful data from a set of occurrences.
Requirement: Optional.
How to use: Provide the subgenus name in parentheses after the genus name.
Examples: *Bombus* (*Cullumanobombus*), *Andrena* (*Melandrena*), *Lasioglossum* (*Dialictus*)
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:subgenus

Term name: **taxonRank**
Definition: The taxonomic rank of the most specific name in the **dwc:scientificName**.
Rationale: Provides important taxonomic context. Accurate identification to the lowest taxonomic rank possible creates the most useful data from a set of occurrences.
Requirement: Optional, but required if **dwc:identificationQualifier** is used.
How to use: Provide the finest resolution of the occurrence species identification.
Examples: genus, species, subspecies
Relevant protocols: All.
For more detail: https://dwc.tdwg.org/terms/#dwc:taxonRank

Term name: **vernacularName**
Definition: A common or vernacular name.
Rationale: Some data collectors, managers, or users may be more familiar with or prefer to use a common name for a species.
Requirement: Optional.
How to use: Provide a common name.

EXAMPLE: Two-spotted bumble bee
RELEVANT PROTOCOLS: All.
FOR MORE DETAIL: https://dwc.tdwg.org/terms/#dwc:vernacularName

APPENDIX 2
Glossary of Terms

ARCHIVE: A data archive can be one of two things: 1) A general data archive for downloading is a snapshot of occurrence data in time, reflecting a dataset at a certain point in history. This type of archive never changes and can be downloaded at any time from a data aggregator, portal, or repository. 2) A Darwin Core archive downloaded from a data aggregator, portal, or repository is a dataset formatted to follow the Darwin Core data standard. A DwC archive can change over time as new records are added by the data provider.

AGGREGATOR: An online source that combines multiple datasets from multiple sources in one place for exploration and download. Examples include GBIF, iDigBio, Symbiota portals, and Discover Life.

CARE: The CARE Principles for Indigenous Data Governance are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination. CARE is an acronym for Collective benefit, Authority to control, Responsibility, and Ethics. (https://www.gida-global.org/care; Carroll *et al*., 2020).

COLLECTIONS MANAGEMENT SYSTEM (CMS): Software used to catalog items in a collection; in natural history, a CMS can be used to organize organismal specimens. Examples include Arctos, Specify, and Symbiota. Data from these systems can be uploaded to the Internet and made openly available.

CONTROLLED VOCABULARY: An accepted list of entries for a type of information (see Term). Controlled vocabulary lists are intended to reduce ambiguity in term use and can be modified if needed.

DARWIN CORE: Darwin Core is a data standard intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, or samples, and related information. (https://www.tdwg.org/standards/dwc/; Wieczorek *et al*., 2012).

DATA AGREEMENT: A contract describing the appropriate use conditions and crediting practices of a dataset. Also referred to as data use or data sharing agreements.

DATA ASSURANCE: A document verifying that a dataset is of high quality and contains accurate information.

DATA ETHICS: Describes proper use and protection of openly accessible data to develop and maintain trust between data generators, managers, and users.

DATA LIFE CYCLE: Eight steps that describe how to work with data throughout and beyond its use in a particular project: Plan, Collect, Assure, Describe, Preserve, Discover, Integrate, and Analyze. (https://escholarship.org/uc/item/7tf5q7n3; Strasser *et al*., 2012).

DATA MODEL: Formal, codified relationships between types of information. When based on an ontology, it is called a semantic model.

Data sovereignty: The idea that data are subject to the laws of the place they were collected.

Data standard: Guidelines to describe and record data (https://www.usgs.gov/data-management/data-standards).

Digital specimen: A digitized record of a physical specimen.

Digital Object Identifier (DOI): A digital identifier of a physical, digital, or abstract object, designed to be both human-usable and machine readable, that allows the persistent, unique identification of and reliable tracking and access to that object (https://www.doi.org/the-identifier/what-is-a-doi/).

Ecological Metadata Language (EML): A metadata standard that describes ecological data (https://eml.ecoinformatics.org/; Jones *et al.*, 2019).

Extended specimen: A digital specimen along with additional ancillary data describing that specimen, including imagery, phenology, molecular information, environmental surroundings, and more (Lendemer *et al.*, 2020).

FAIR: Describes data that is Findable, Accessible, Interoperable, and Reproducible (https://www.go-fair.org/fair-principles/; Wilkinson *et al.*, 2016).

Identifiers: Machine-readable: Identifiers meant for use by computer software. Universally unique: A machine-readable identifier consisting of a 128-bit label that should only ever be created once. Human-usable: Identifiers that people use to manage data. Unique: a human-usable, usually human-created identifier for a specific purpose.

Integrated Publishing Toolkit (IPT): Software used to publish data to GBIF (https://www.gbif.org/ipt; Robertson *et al.*, 2014).

Key:value pair: A means of providing complex information in a machine-readable manner. A key is the type of information being provided, and the value is the actual information.

License: A document that describes conditions of acceptable use of openly shared data. Assigning a license to a dataset promotes proper crediting. Creative Commons licenses are commonly used; beyond the public domain license, there are six other licenses that can be applied to occurrence datasets: https://creativecommons.org/share-your-work/cclicenses/

Mapping: Translating data column headings to Darwin Core terms.

Metadata: Data that describes other data, including authorship, date created, date published, and license restrictions.

Occurrence data: Records of organisms that include the date collected, collector identification, geographic location, and other ancillary information.

Openly available: Data is considered openly available if it can be downloaded freely without explicit permission from the provider (*i.e.*, no email exchange is required to obtain a dataset) and the conditions for using the data are made clear by the provider. Use conditions are typically provided by applying a license (see License).

Ontology: A description of how various concepts within an area of information relate to each other. An ontology can be used to create a data model.

ORCID iD: A unique identifier used by scientists to link their output across the Internet, obtained and recorded at the Open Research and Contributor ID registry (https://orcid.org/; Haak *et al*., 2012).

PORTAL: A place where data or datasets can be downloaded. Portals provide downloadable access to data.

RECORD: A row of data in a digitized file.

REPOSITORY: A place where data are stored but not aggregated. Datasets can be downloaded individually from individual providers. Repositories can serve as an archive, and archived data are never changed. Examples: Zenodo (https://zenodo.org/), FigShare (https://figshare.com/), the Environmental Data Initiative repository (EDI; https://portal.edirepository.org/nis/home.jsp), and Dryad (https://datadryad.org/stash).

THEMATIC COLLECTIONS NETWORK (TCN): A Thematic Collections Network is a network of institutions with a strategy for digitizing information that addresses a particular research theme. Once digitized, data are easily accessed and available for other research and educational use. (https://www.idigbio.org/content/thematic-collections-networks). Current TCNs that are digitizing information about wild bees are iDigBees (DBI#2216927; https://idigbees.org) and Big-Bee (DBI#2102006; https://big-bee.net).

TDWG: Once known as the Taxonomic Databases Working Group, the Biodiversity Information Standards (TDWG) works to establish international collaboration among the creators, managers and users of biodiversity information and to promote the wider and more effective dissemination and sharing of knowledge about the world's heritage of biological organisms. https://www.tdwg.org/

TERM: In Darwin Core, a term is a label for a type of information and is used as a column heading in a digitized file of occurrence data. The full list of Darwin Core terms is found at https://dwc.tdwg.org/terms/

APPENDIX 3
Additional Resources

Native Bee Monitoring Rcn Data Management Workshop:

- Agenda: https://www.nativebeemonitoring.org/news/workshop-data-management
- Videos:
  - Workshop presentations Day 1: https://youtube.com/playlist?list=PLh3NEUAQ4ng7eQF_xnDYgreNjazdNXiQG&si=bhsF6MhMPKIu7fkw
  - Workshop presentations Day 2: https://youtube.com/playlist?list=PLh3NEUAQ4ng63v-hUamq1lEpLaUvg2JAF&si=vEvF1NYL6fIUaPCP
  - Data management and digitization: examples and resources: https://youtube.com/playlist?list=PLh3NEUAQ4ng5IgHQHZnftYtNZzPzdz0gr&si=3U_VSxfYdbj92gdl
  - Data management with different types of data: https://youtube.com/playlist?list=PLh3NEUAQ4ng6GXVYe2xJHQVVvWkbglVdZ&si=bRRe8UGHCllIF5NV

Links Referenced In This Manuscript:

- Darwin Core term list: https://dwc.tdwg.org/terms/
- GBIF: https://www.gbif.org/
  - Citation guidelines: https://www.gbif.org/citation-guidelines
  - Global Registry of Scientific Collections (GrSciColl): https://scientific-collections.gbif.org/
  - GBIF North America: https://www.gbif-north-america.org/community/
  - Templates to map occurrence data to Darwin Core terms:
- For occurrence data: https://ipt.gbif.org/manual/en/ipt/latest/occurrence-data#templates
- For sampling event data: https://ipt.gbif.org/manual/en/ipt/latest/sampling-event-data#templates
- Symbiota: https://symbiota.org/
  - Bee Library: https://library.big-bee.net/portal/
  - SCAN: https://scan-bugs.org/portal/
  - Ecdysis: https://ecdysis.org/index.php
  - Symbiota Documentation: https://biokic.github.io/symbiota-docs/
- Citation guidelines: https://biokic.github.io/symbiota-docs/coll_manager/citation/

- BeeBDC:
  - Homepage: https://jbdorey.github.io/BeeBDC/index.html
  - GitHub Repository: https://github.com/jbdorey/BeeBDC
- Codes of Conduct or Norms for Data Use:
  - USGS Bird Banding Laboratory Data Release Policy: https://www.usgs.gov/labs/bird-banding-laboratory/science/data-release-policy
  - CARE Principles for Indigenous Data Governance (Carroll *et al.*, 2020): https://www.gida-global.org/care
  - US Federal Data Strategy Data Ethics Framework: https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf
  - Canadensys Norms for Data Use: https://github.com/Canadensys/norms-for-data-use
  - VertNet Norms for Data Use: https://vertnet.org/resources/norms.html
- Creative Commons: https://creativecommons.org/share-your-work/cclicenses/
- UUID generator tools:
  - UUID R Package: https://www.rforge.net/uuid/, https://cran.r-project.org/web/packages/uuid/uuid.pdf
  - UUID Generator web application: https://www.uuidgenerator.net/
  - Guidance from iDigBio on generating UUIDs in Excel: https://www.idigbio.org/wiki/images/0/03/GUIDgeneration.pdf

OTHER RELEVANT DATA STANDARDS: https://www.gbif.org/standards

- Ecological Metadata Language (EML): https://eml.ecoinformatics.org/
- Developing standards
  - Humboldt Extension to Darwin Core to incorporate inventory and monitoring information, including sampling protocol and effort: https://eco.tdwg.org/
  - Plant-pollinator terms
- REBIPP: https://ppi.rebipp.org.br/
- Term list: https://ppi.rebipp.org.br/terms/
- WorldFAIR: https://worldfair-project.eu/agricultural-biodiversity/

OTHER TOOLS OR DATA SHARING PLATFORMS

- iDigBio: https://www.idigbio.org/
- USDA Plants: https://plants.usda.gov/home
- Discover Life: https://www.discoverlife.org/
- Global Biotic Interactions (GloBI): https://www.globalbioticinteractions.org/

Other Options For Collections Management Systems:

- Specify: https://www.specifysoftware.org/
- Arctos: https://arctosdb.org/

Learning Opportunities:

- iDigBio Digitization Academy: https://digitizationacademy.org/
- TDWG YouTube channel: https://www.youtube.com/@tdwg/featured
- Environmental Data Initiative: https://edirepository.org/
- Entomological Collections Network listserv https://ecnweb.net/resources/listserv/

Further Reading On Sharing And Using Occurrence Data:

Ball-Damerow, J.E., L. Brenskelle, N. Barve, P.S. Soltis, P. Sierwald, R. Bieler, R. LaFrance, A.H. Ariño, & R.P. Guralnick. 2019. Research applications of primary biodiversity databases in the digital age. *PLoS ONE* 14(9): e0215794. https://doi.org/10.1371/journal.pone.0215794

Costello, M.J. 2009. Motivating online publication of data. *BioScience* 59(5): 418–427. https://doi.org/10.1525/bio.2009.59.5.9

Costello, M.J., W.K. Michener, M. Gahegan, Z.Q. Zhang, & P.E. Bourne. 2013. Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution* 28(8): 454–461. https://doi.org/10.1016/j.tree.2013.05.002

Costello, M.J., & J. Wieczorek. 2014. Best practice for biodiversity data management and publication. *Biological Conservation* 173: 68–73. https://doi.org/10.1016/j.biocon.2013.10.018

Costello, M.J., B. Vanhoorne, & W. Appeltans. 2015. Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology* 29(4): 1094–1099. https://doi.org/10.1111/cobi.12496

Guralnick, R.P., A.W. Hill, & M. Lane. 2007. Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10: 663–672. https://doi.org/10.1111/j.1461-0248.2007.01063.x

Heberling, J.M., J.T. Miller, D. Noesgaard, & D. Schigel. 2021. Data integration enables global biodiversity synthesis. *PNAS* 118(6): e2018093118 https://doi.org/10.1073/pnas.2018093118

Jetz, W., M.A. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M.J. Costello, M. Fernandez, G.N. Geller, P. Keil, C. Merow, C. Meyer, F.E. Muller-Karger, H.M. Pereira, E.C. Regan, D.S. Schmeller, & E. Turak. 2019. Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution* 3: 539–551 https://doi.org/10.1038/s41559-019-0826-1

Journal of Melittology

A Journal of Bee Biology, Ecology, Evolution, & Systematics

The *Journal of Melittology* is an international, open access journal that seeks to rapidly disseminate the results of research conducted on bees (Apoidea: Anthophila) in their broadest sense. Our mission is to promote the understanding and conservation of wild and managed bees and to facilitate communication and collaboration among researchers and the public worldwide. The *Journal* covers all aspects of bee research including but not limited to: anatomy, behavioral ecology, biodiversity, biogeography, chemical ecology, comparative morphology, conservation, cultural aspects, cytogenetics, ecology, ethnobiology, history, identification (keys), invasion ecology, management, melittopalynology, molecular ecology, neurobiology, occurrence data, paleontology, parasitism, phenology, phylogeny, physiology, pollination biology, sociobiology, systematics, and taxonomy.

The *Journal of Melittology* was established at the University of Kansas through the efforts of Michael S. Engel, Victor H. Gonzalez, Ismael A. Hinojosa-Díaz, and Charles D. Michener in 2013 and each article is published as its own number, with issues appearing online as soon as they are ready. Papers are composed using Microsoft Word® and Adobe InDesign® in Lawrence, Kansas, USA.

Journal of Melittology is registered in ZooBank (www.zoobank.org), and archived at the University of Kansas and in Portico (www.portico.org).

http://journals.ku.edu/melittology
ISSN 2325-4467