



# Table of Contents

Colophon .....	1
Suggested citation .....	1
Authors .....	1
Licence .....	1
Acknowledgement .....	1
Persistent URI .....	1
Document control .....	1
Cover image .....	1
1. Introduction .....	2
1.1. Scope .....	2
1.2. Target audience .....	3
1.3. Using this guide .....	3
1.3.1. Data mapping template .....	3
1.3.2. Example data .....	4
2. Biological survey data .....	5
2.1. Making biological survey and monitoring data FAIR and open .....	6
2.2. Darwin Core standard and biological survey and monitoring data .....	7
2.3. Biological survey and monitoring data in GBIF: Darwin Core Archives (DwC-A) .....	7
3. Mapping survey and monitoring data to Darwin Core .....	10
3.1. Translating survey design into Darwin Core Event structure .....	11
3.2. Non-nested datasets .....	11
3.3. Nested datasets .....	11
3.3.1. Simple nested data structures .....	12
3.3.2. Simple nested datasets with Project-level information .....	13
3.3.3. Deeply nested datasets .....	14
3.3.4. Constructing a dataset schematic .....	16
4. Resource metadata .....	19
5. Survey Event data: capturing the context of biological surveys and monitoring data .....	20
5.1. Survey design .....	21
5.2. Project information .....	22
5.3. Survey site .....	23
5.3.1. Site description .....	23
5.3.2. Site locality .....	24
5.3.3. Survey site area .....	24
5.3.4. Additional survey site information .....	25
5.3.5. Vegetation cover .....	26
5.4. Survey date and time .....	26
5.5. Methodology or sampling protocol .....	27
5.5.1. Event type .....	27
5.5.2. Sampling protocol .....	29
5.5.3. Absences .....	30
5.5.4. Abundance .....	31
5.5.5. Material samples .....	31

5.5.6. Vouchers .....	31
5.5.7. Sensitive data: data generalization & information withheld .....	32
5.5.8. Least specific target category quantity inclusive .....	32
5.5.9. Verbatim fields .....	33
5.6. Scope and completeness .....	33
5.6.1. Verbatim scope .....	34
5.6.2. Taxonomic scope .....	34
5.6.3. Organismal scopes .....	35
5.6.4. Bycatch .....	36
5.6.5. Habitat scope .....	37
5.7. Sampling Effort .....	37
6. Mapping additional survey event information: DwC Extensions .....	39
6.1. Extended measurement or fact (eMoF) extension .....	39
6.2. Relevé extension .....	39
7. Mapping Occurrence information .....	40
7.1. Reporting multiple individuals as a single Occurrence .....	40
7.2. Reporting multiple individuals as multiple occurrences .....	40
7.3. Reporting absences .....	41
7.4. Reporting abundances .....	43
7.5. Capturing species co-occurrence and species interactions .....	44
8. Specific biological survey types .....	46
8.1. Camera trap survey data .....	46
8.2. DNA and metabarcoding data .....	46
8.3. Environmental impact assessments .....	46
8.4. Freshwater biodiversity data .....	46
8.5. Vector-pathogen data .....	46
8.6. Private companies .....	46
9. Additional guidance and reaching out for assistance .....	47
10. Feedback .....	47
Glossary .....	48
References .....	51

# Colophon

## Suggested citation

Ingenloff K, Svenningsen C, Earl C, Shimabukuro PHF, Sica Y, Gan Y-M, Kachian ZR, Brenton P, Hochachka W, Wieczorek J, Stevenson R, Kazem A, Baskauf S, Zermoglio PF, Bloom D, Rodrigues A, Gamboa Martínez J & Schigel D. Guide for publishing biological survey and monitoring data to GBIF. GBIF Secretariat: Copenhagen. <https://doi.org/10.35035/doc-ynvs-eh84>

## Authors

Ingenloff K, Svenningsen C, Earl C, Shimabukuro PHF, Sica Y, Gan Y-M, Kachian ZR, Brenton P, Hochachka W, Wieczorek J, Stevenson R, Kazem A, Baskauf S, Zermoglio PF, Bloom D, Rodrigues A, Gamboa Martínez J & Schigel D

## Licence

The document *Guide for publishing biological survey and monitoring data to GBIF* is licensed under [Creative Commons Attribution-ShareAlike 4.0 Unported License](#).

## Acknowledgement

This guide was produced under [the BioDT project](#), which received funding from the European Union's Horizon Europe research and innovation programme under [grant agreement No 101057437](#).

## Persistent URI

<https://doi.org/10.35035/doc-ynvs-eh84>

## Document control

v1.0.2, published 2025-09-24.

## Cover image

Illustration by Javier Gamboa, GBIF secretariat 2025, licensed under [CC BY](#)

# 1. Introduction

**Biological surveys**, systematic efforts to collect information about the biological organisms of a specific area at a given time, are critical to helping us understand and monitor changes in our environment. Also referred to as biodiversity surveys, these efforts employ a wide variety of methods or protocols to contribute to our knowledge about species distributions and abundances, community composition, and ecological relationships. Different communities also refer to biological surveys as ecological inventories, biodiversity monitoring, biological sampling or recording, among other terminology; we will use these terms interchangeably in this guide, and will often simply refer to them as 'surveys.' Biodiversity surveys support larger **ecological monitoring** efforts aimed at evaluating ecosystem health and ecological response to climate change, supporting conservation efforts, informing policy and management, and improving public awareness and education about the values of biodiversity. These monitoring efforts can be question-driven, with protocols designed to answer a particular question or series of questions; to emphasize general monitoring, focused on establishing a baseline and building a record; or take more of a 'naturalist' approach, with repeated data collection occurring out of curiosity. Most commonly, at the moment of writing, monitoring language appears in the context of environmental monitoring or science policy at the province, city, state, or state level.

Governing body and international organizational reports consistently emphasize that available data are scarce for a proper assessment of nearly all facets of biodiversity in response to the current global biodiversity crisis [IPBES 2019]. One way to address the need for extensive biodiversity data is to aggregate existing datasets from prior and disparate biological surveys, monitoring efforts, and data catalogues. International organizations (e.g., **Convention on Biological Diversity**, **Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services**), governmental or regional organizations (such as the Australian National Data Service and the EU's Open Data Directive), funding agencies (e.g., EU Horizon Europe, **US National Science Foundation**), and scientific journals (e.g., **PLOS ONE**, **Pensoft**, **Nature Research Journals**) are increasingly adopting the requirement that biodiversity data be made FAIR (findable, accessible, interoperable, reusable) and open [Wilkinson et al. 2016]. These mandates are designed to enhance transparency, reproducibility, and the collective impact of biodiversity research and conservation efforts. Although biodiversity data are increasingly findable and accessible thanks to initiatives and mandates requiring research data and outputs to be made FAIR, it is often still difficult to assess the usability of data for aggregation or application in larger analyses due to lack of standardization.

GBIF, the **Global Biodiversity Information Facility**, is among the leading open access FAIR biodiversity data infrastructures. In 2025, users can access more than 3 billion species occurrence records from approximately 2,300 publishing institutions globally; but, it is difficult to assess fitness for use of these data in analyses requiring integration of biodiversity survey data. Data shared through GBIF are standardized using the **Darwin Core (DwC)** data standard, managed by **Biodiversity Information Standards (TDWG)**, to facilitate data discovery and support easier aggregation of datasets. Recent improvements in DwC provide a means by which to capture the structural and methodological complexities of biodiversity surveys (see the **Humboldt extension for ecological inventories**), which facilitates efforts to identify appropriate datasets and aggregate data from heterogeneous sources.

This guide serves as a tool to help holders of biological survey and monitoring data capture key facets of survey design using the Darwin Core standard to facilitate FAIR and open sharing of their data through GBIF.

## 1.1. Scope

This guide aims to help those with biodiversity survey and/or monitoring data improve the interoperability of their data, thus facilitating increased data reuse, through application of the **Darwin Core Biodiversity Data Standard**. This guide provides an overview of the primary components of

biodiversity survey data in the context of the Darwin Core standard, [DwC Events](#), and the [Humboldt extension for ecological inventories](#). In particular, this guide assists the reader in structuring their data as a Darwin Core Archive and walks the reader through the process of mapping their data to DwC terms. Readers will be pointed to existing additional documentation where available.

## 1.2. Target audience

This guide aims to help ecologists, researchers, and data managers from any organization or group (be they commercial, government agencies, non-governmental organizations, research groups, private sector, or other) wanting to standardize and share biodiversity survey and monitoring data, specifically those aiming to format their data with the intent of publishing these data to [GBIF](#).

If you are already comfortable with Darwin Core Event datasets and are simply seeking guidance in applying the Humboldt extension, refer to the [Survey and monitoring data quick start guide \[Ingenloff 2025\]](#).

## 1.3. Using this guide

Throughout the guide, Darwin Core terms will be written in fixed width font and preceded by their namespace abbreviation and a colon ('dwc:' or 'eco:') to denote the DwC core or extension to which the term belongs. For example, the Darwin Core event ID term will appear as `dwc:eventID` and the Humboldt Extension verbatim target scope term will be written as `eco:verbatimTargetScope`.

Terms are linked with their respective term [internationalized resource identifier](#) (IRI, e.g., `eco:protocolNames`).

*Namespace abbreviations and usage examples for DwC terms*

Namespace abbreviation	Core or extension name	Example
dwc	Darwin Core  (applies to Event core, occurrence extension, extended measurement or fact extension, and related resource extension terms)	<code>dwc:eventID</code>
eco	Humboldt extension for ecological inventories	<code>eco:verbatimTargetScope</code>

### Term usage recommendations

Each term mentioned in this guide is associated with one of 3 usage recommendations.

- **Required** terms must be populated and included with a dataset for publication to GBIF or for reusability.
- **Recommended** terms enhance the value and broader usefulness of a dataset with improved information about event localities, sampling context, methods, and/or scopes.
- **Share if available** terms can further enhance the potential usefulness of a dataset.

### 1.3.1. Data mapping template

A basic data template is available to facilitate mapping and preparation of biodiversity survey and monitoring data for formatting as a Darwin Core Archive. The template can be accessed as a single .xlsx file or as three separate .csv files.

- Survey data template (.xlsx)
- Survey event table template (.csv)
- Survey occurrence table template (.csv)
- Survey template README (.csv)

Table	Description
event	<p>Terms in the <b>event</b> table are used to capture survey Event information (i.e. information that applies to the observations of all taxa) including survey design, protocol(s), scopes, and effort. Terms from the <b>Darwin Core Event class</b> and the <b>Humboldt extension</b> are included.</p> <p>Column heads are populated with the DwC Event core and Humboldt extension terms referenced in this guide. The rows beneath each term include term definitions, comments, recommended usage for publication in GBIF, and additional comments or usage guidance.</p>
occurrence	<p>Terms in the <b>occurrence</b> table should be used to capture information about the occurrence of a single taxon with terms from the <b>Darwin Core Occurrence extension</b>.</p> <p>Column heads are populated with the DwC Occurrence extension terms referenced in this guide. The rows beneath each term include term definitions, comments, and recommended usage for publication in GBIF. Additional Occurrence extension terms should be added to your own data <b>occurrence</b> table as appropriate for your dataset.</p>
README	The README table provides additional information about the structure and information included in each data table.

### 1.3.2. Example data

The authors are collaborating with the **National Ecological Observatory Network (NEON)** to develop a comprehensive example dataset to accompany this guide. The guide will be updated as soon as the dataset is available.

The GBIF datasets listed below implement some Humboldt extension terms and may serve as useful references on Event dataset structure and term usage.

- Faveyts W and Cooleman S (2025). Bird census counts at the Zwin Nature Park. Version 1.5. Belgian Biodiversity Platform. Sampling event dataset <https://doi.org/10.15468/saesvn>.
- Palpurina S (2025). Vegetation plots collected in dry grasslands throughout Bulgaria and Romanian Dobrudzha. Version 1.12. Masaryk University, Department of Botany and Zoology. Sampling event dataset <https://doi.org/10.15468/pkx4tg>.
- Piesschaert F, Vermeersch G, Brosens D, Westra T, Desmet P, Feys S, Van de Poel S, Pollet M, and Cooleman S (2025). ABV - Common breeding birds in Flanders, Belgium (post 2016). Version 1.14. Research Institute for Nature and Forest (INBO). Sampling event dataset <https://doi.org/10.15468/pj2v6h>.
- van Klink R and Gerrits G (2025). Biological Station Wijster standard trapping program: Sampling event data for ground beetles (Coleoptera: Carabidae). Version 1.3. WBBS foundation. Sampling event dataset <https://doi.org/10.15468/3mcqja>.



## 2. Biological survey data

**Biological (a.k.a. biodiversity) surveys** aim to identify and document the presence (and often quantify the abundance) of a particular group of organisms (taxonomic scope) in a specific location or series of locations (spatial scope) over a defined period (temporal scope) using an explicit methodological approach (protocols, sampling design). A simple biological survey may take place at a single location or site, implementing a single sampling protocol, and occurring at a single time with no repeated visits to the survey site. More complex surveys may take place at multiple sites, employ a broad suite of methods, including field observations, sampling techniques, deployment of camera traps, acoustic monitoring, genetic analysis, and remote sensing, with one or more repeat visits to some or all of the surveyed sites (e.g., time series data). As such, biological survey and monitoring data typically need to include a wide range of information to comprehensively document the methods implemented, and recorded presence, abundance, and condition of species and their traits and habitats. Incidental or opportunistically collected data are not considered survey data.

The details about a survey (how it was carried out, the spatio-temporal scope, the taxonomic group targeted, who was involved, etc.) are critical to properly understanding the structure of the data resulting from the survey and how it can be analyzed, (re-)interpreted, and (re-)used for other purposes. Despite its inherent value, this detailed information is often treated as metadata and captured in an unstructured manner that makes it nearly impossible to take full advantage of the breadth of information available. Standardizing the way this information is reported provides a means of understanding and interpreting a dataset without requiring the intimate knowledge of a dataset owner or creator (on recontextualization of data see [Leonelli 2016] page 32).

The breadth of information that can be captured from structured reporting of biological survey design alongside the actual data recorded during a survey includes:

- **Survey structure:** Survey structure includes information about the study area and sampling units of a survey. It provides a means by which to understand how data collected during a survey relate to each other in location, scope, and sampling date and time.
- **Survey methods:** Survey methods includes detailed information about the sampling protocol implemented (e.g., protocol name, relevant references, details of techniques implemented and equipment used) and the type(s) of data collected.
- **Survey scope:** Scopes define the overall objectives of a survey and will vary depending on the purpose of the survey. Common scope types include:
  - **Spatial scope:** Spatial scope refers to the geographic area of interest of a survey. It can include information about the location of each survey site including geographic coordinates with geodetic datum, site description (locality name, habitat type, microhabitat), and environmental data (e.g., physical parameters, vegetation, water quality). It can also identify any areas or habitats specifically targeted for or excluded from survey efforts.
  - **Temporal scope:** Temporal scope identifies the time during which a survey took place (e.g., a single day, single season, multiple seasons).
  - **Taxonomic scope:** Taxonomic scope identifies any group(s) of organisms specifically targeted for, or excluded from, a survey.
  - **Organismal scope:** Organismal scope identifies the type(s) of organisms specifically targeted for, or excluded from, a survey. Organismal scope may include age, sex, life stage, reproductive status, etc.
- **Survey effort:** Survey effort defines the amount of effort put into conducting a survey (for example, the number of trap nights per sample site) and describes any protocol used to assess effort.



At a higher level of aggregation, **compilations** are a type of biological survey which results from combining existing surveys, rather than generated *de novo* from observations or samples (see `eco:inventoryTypes` and `eco:compilationTypes`). Compilations may aggregate surveys using multiple protocols, processes, and observers, or other compiled data sources and literature searches. They are typically combinations of multiple broad studies performed within a broad spatial scope (e.g. [Dimaki & Legakis 1999]).

## 2.1. Making biological survey and monitoring data FAIR and open

Making biological survey and monitoring data **FAIR** and **open** enhances scientific research and collaboration, enables large-scale analyses through data aggregation, improves data quality, fosters innovation, and promotes efficient use of resources.

*Guiding principles for making biological survey and monitoring data FAIR and open*

Findable	<ul style="list-style-type: none"> <li>• Metadata: Create comprehensive metadata describing the dataset using a standardized schema such as Ecological Metadata Language (EML).</li> <li>• Persistent identifiers: Assign persistent identifiers (e.g., Digital Object Identifiers or DOIs) to datasets to ensure they can always be found.</li> <li>• Repositories: Register datasets with reputable data repositories that support metadata standards and publish the dataset through a data aggregator such as GBIF to enhance discoverability.</li> </ul>
Accessible	<ul style="list-style-type: none"> <li>• Open licensing: Apply open licenses such as Creative Commons (CC0, CC-BY) to allow others to freely access and use the data.</li> <li>• Public repositories: Ensure that your chosen data repository is publicly accessible and does not require subscriptions or membership.</li> <li>• Data availability statement: Include data availability statements in publications with links to the repositories where data are stored. Be aware of firewalls and other forms of access restrictions unless justified (see [Chapman 2020]).</li> </ul>
Interoperable	<ul style="list-style-type: none"> <li>• Standard formats: Use standard, non-proprietary data formats (e.g., CSV, JSON, XML) to ensure broad compatibility with software tools.</li> <li>• Data standards: Adhere to relevant data standards and schemas, such as Darwin Core, to facilitate aggregation and integration with other datasets.</li> <li>• Controlled vocabularies: Use controlled vocabularies and ontologies to ensure consistent terminology and semantic clarity.</li> </ul>
Reusable	<ul style="list-style-type: none"> <li>• Data reuse terms: Clearly state the terms of use through licenses like Creative Commons to ensure users understand how they can reuse the data legally.</li> <li>• Data documentation: Include detailed documentation covering data collection methods, data processing steps, and any limitations or uncertainties in the data.</li> <li>• Data provenance: Document the origin and lineage of the data to provide context and credibility. Rely on DOIs for data citation needs.</li> <li>• Validation and cleaning: Ensure data quality by performing validation, error checking, and data cleaning. Document these processes to maintain transparency.</li> </ul>

## 2.2. Darwin Core standard and biological survey and monitoring data

The [Darwin Core \(DwC\) biodiversity data standard](#) is a community-maintained biodiversity information standard. The primary goal of DwC is to support biodiversity informatics by making data interoperable and reusable across myriad platforms and applications. DwC provides a set of terms, definitions, and guidelines designed to facilitate the exchange of biological data. DwC **terms** are used to describe and share biodiversity data. Each term has an accepted definition accompanied by comment(s), usage examples (for example, see `eco:[eco:verbatimTargetScope]` and `eco:[eco:protocolNames]`), and in some cases are based on or recommend use of a controlled vocabulary (a list of accepted values that can be used for the term). The process of matching data or information from one dataset to the terms of another such as DwC is referred to as [mapping](#).

Implementation of the DwC standard reduces errors and inconsistencies in data, and enhances data discoverability, which ultimately facilitates data reuse. DwC includes terms for describing species occurrences and biodiversity surveys, including terms for methodology, survey location (site), survey date(s), taxonomy, and other relevant attributes. DwC **extensions** provide additional terms and properties for specific types of biodiversity data enabling researchers to capture a broader range of information tailored to particular needs, such as data on ecological interactions, genetic sequences, or sampling events.

In a Darwin Core context, biological survey and monitoring data are best captured as [Events](#), where time- and space-specific detection are documented centrally and separately from the list of species recorded in each Event. Historically, DwC evolved from natural history collections to other biodiversity data contexts, and until recently struggled to effectively capture more complex data like biological surveys. Specifically, detailed information about survey design, sampling methods and protocols, scope, and completeness were captured in an unstructured manner relegated largely to verbatim text fields such as `dwc:samplingProtocol` and `dwc:samplingEffort`. The [Humboldt Extension for Biological Inventories](#) (HE), an extension to the DwC Event core, provides data publishers with a means by which to share biological survey and monitoring data in a structured manner to increase the findability of datasets and improve the chance of dataset reuse. The extension added 55 terms to the DwC Event class vocabulary by which to capture components of the contextual information about a survey previously lost as unstructured metadata (see the terms list in [\[TDWG Humboldt Extension Task Group 2024\]](#)).

## 2.3. Biological survey and monitoring data in GBIF: Darwin Core Archives (DwC-A)

Biodiversity data can be shared to GBIF in multiple ways; however, data need to be shaped to conform to the current data model which is structured around [Darwin Core Archives \(DwC-A\)](#). Data published to GBIF are shared as one of four **dataset categories**:

- [metadata-only](#),
- [taxonomic checklists](#),
- [occurrences](#), or
- [sampling events](#).

These categories are each associated with a 'core' ([Taxon](#), [Occurrence](#), [Event](#)) which defines how the data should be formatted. Each core can be supplemented with one or more [GBIF registered extensions](#).

In GBIF, biological survey and monitoring data are broadly referred to as **sampling Event data** and should be formatted using the **DwC Event core**. DwC Event data have been publishable through GBIF since 2016; as of 2025, more than 4,000 Event datasets are discoverable. To publish an Event core dataset to GBIF, the dataset must be structured as a **Darwin Core Archive (DwC-A)** consisting of the following files (see also Figure 1 in [GBIF 2018]):

- **Metafile:** **Required** The metafile describes what files exist in the DwC-A and how the columns in each data file map to Darwin Core terms. The metafile is essentially a resource map.
- **Resource metadata:** **Required** The resource metadata file describes the dataset context in more detail e.g., description of the dataset, people involved, etc. using terms derived from Ecological Metadata Language (EML).
- **Event core:** **Required** The Event table(s) includes DwC Event and Humboldt extension terms describing survey-level information (e.g., protocol, survey scope, sampling effort and completeness).
- **Occurrence extension:** **Optional** The occurrence extension file(s) to an Event core dataset contains associated organismal Occurrence information.
- **DwC extension file(s):** **Optional** Additional tables may contain data that further expands on details relating to the survey (see below for more information about extensions). See the table below for an overview of GBIF registered extensions.

*A non-exhaustive list of GBIF registered DwC extensions that can and cannot currently be published with a Darwin Core Event dataset to GBIF.*

DwC extensions that can currently be published through GBIF with a DwC Event core dataset	<ul style="list-style-type: none"> <li>• <b>Occurrence</b> - captures the content of organismal Occurrence records</li> <li>• <b>Humboldt extension</b> - provides extended support for data coming from biodiversity surveys</li> <li>• <b>Extended measurement or fact (emof)</b> - extends the generic measurements or facts terms. When used with Event core it allows to create an additional link between the emof and the Occurrence extensions.</li> <li>• <b>Resource relationship</b> - extended support for reporting relationships between the core dataset and extensions or external data)</li> <li>• <b>Relevé</b> - supports vegetation plot survey (relevé) measurements</li> <li>• <b>Media</b> - supports metadata from biodiversity multimedia resources and collections applied to an Event</li> </ul>
DwC extensions that <b>cannot</b> currently be published through GBIF with a DwC Event core dataset	<ul style="list-style-type: none"> <li>• <b>DNA-derived data</b> - captures information relating to DNA-derived data. GBIF currently recommends the extension is used with Occurrence core to be able to capture sequence/Occurrence specific information.</li> <li>• <b>Media</b> - supports metadata from biodiversity multimedia resources and collections applied to Occurrences</li> <li>• <b>Chronometric age</b> - captures true age information of a historic specimen collected long after the dwc:organism was deceased</li> <li>• <b>Identification history</b> - specific for Occurrences and used as an extended support to capture multiple identifications of the same dwc:Organism</li> </ul>

A Darwin Core Archive for biodiversity survey and monitoring data will require at least two tables: **metadata** and **event**. The DwC-A will have an additional table for each extension (e.g., Occurrence, extended measurement or fact) included with the archive.

## A note on extensions and DwC-A in GBIF

The current GBIF data model permits only a single 'layer' of extensions (that is, an extension cannot be attached to an extension). This means that an Event core dataset may include the Occurrence extension, as well as other relevant extensions, but the DwC-A could not support an extension connected to the Occurrence extension.

Furthermore, some extensions can only be used in combination with a specific core while others can be used with multiple cores. For example, the [DNA-derived data extension](#) can currently only be used in conjunction with an Occurrence core dataset event though it was developed for use with both Occurrence and Event cores, and the [Humboldt extension](#) can only be used with the Event core.

This limitation applies only to publishing the data to GBIF. Follow GBIF's work on the evolving [data model](#).

### 3. Mapping survey and monitoring data to Darwin Core

Data standardization is often wrongly perceived as an invasion of an established or bespoke data collection system. In reality, data standardization is simply a transformation of the data export while the source data system remains intact. The following sections will guide you through the process of mapping the Event-level (sampling context) information of your biodiversity survey and/or monitoring data to the Darwin Core data standard.

In practice, the process of mapping survey data to DwC for publication in GBIF will roughly follow these steps:

- **Identification of the structure, or hierarchy, of the data:** In essence, this is the process of translating the sampling design of a biological survey (or series of surveys) to Darwin Core Event format. Does the dataset consist of a single survey at a single location? Multiple surveys conducted at different times at the same location? Or a series of surveys at different locations? See [Translating survey design to DwC Event data structure](#).
- **Identification of the data composition and DwC vocabulary needs:** Before actually mapping data to terms, it is useful to identify the vocabulary extensions that will be necessary to report all data (or as much data as possible) from the dataset. Available extensions can be explored via the [GBIF registered extensions](#) and [TDWG biodiversity information standards](#). See [Constructing a dataset schematic](#).
- **Mapping of survey (Event) information to DwC Event terms:** Information about each biological survey (simply referred to as an 'Event' or 'sampling Event') will be mapped to [DwC Event class](#) and [Humboldt extension](#) terms and saved in an [event](#) table or tables. Event-level data include the contextual information that applies to all Occurrence and ancillary data collected or recorded during an Event. Examples include information about the survey design, site (e.g., location, date), protocol(s), scope(s), and sampling effort. Resource: see the `data/event_template_wHE_event-table.csv['event' table in the data mapping template]`. See [Survey Event data: capturing the context of biological survey and monitoring data](#).
- **Mapping of Occurrence data to the DwC Occurrence extension:** Organism Occurrence information collected during biological surveys (e.g., scientific name, additional organismal information) will be shared in an independent 'occurrence' table using the [Occurrence extension](#). See the `data/event_template_wHE_occurrence-table.csv[occurrence table]` in the data mapping template and [Mapping Occurrence information](#).
- **Mapping of ancillary data to appropriate extensions:** Additional information collected during a survey that requires use of one or more extensions should be mapped so as to link the information to the appropriate Event(s) or organisms via the relevant Event identifiers.

**The recommended best practice is to map as much of your data as possible using all existing vocabulary standards and extensions necessary for your data.**

The landscape of biodiversity data in GBIF is always evolving. While some data cannot yet be published to GBIF with a DwC Event dataset, GBIF maintains stepwise efforts to improve the underlying data model and expand the breadth of data types and complexity that can be accommodated. Data that cannot be published now will likely be publishable in the future. As such, mapping as much data in a dataset as possible now reduces the amount of time and energy spent overall, removing the need to revisit the process at a later date.

## 3.1. Translating survey design into Darwin Core Event structure

**Biological survey design**, the sampling structure of a biological survey, varies widely. Identifying how to best translate survey design to DwC Event core is the most difficult part of mapping a survey dataset. DwC defines an **Event** as 'an action that occurs at some location during some time', such as a specimen collection expedition, a camera trap image capture, or a marine trawl. This broad definition of Event means biological surveys can be framed as a single Event or as a series of Events nested within Events using a **parent-child** relationship as necessary. The **sampling Event hierarchy** is the translation of survey design into an Event-based perspective using Darwin Core.

Sharing biodiversity data in a way that clearly and accurately reflects survey design helps ensure accurate understanding and interpretation of the information contained in a dataset enabling potential data users to more readily assess the appropriateness of the data for inclusion in their own analyses.

## 3.2. Non-nested datasets

**Non-nested datasets** reflect a simple or flat survey design structure (Figure 1). These are typically simple datasets consisting of:

- a single sampling Event occurring at a particular place and time and conducted using a single standardized sampling protocol that is not repeated and is not necessarily part of a larger sampling schema (Figure 1a), or
- a series of single sampling Events that are not joined by a larger parent Event (Figure 1b). A compilation (e.g., a combination of unrelated surveys, compiled data sources and/or literature searches, see the **Biological survey data** section) could be a special case of non-nested dataset where there is a unique Event level that describes the compilation itself (e.g., the broad area where multiple surveys are aggregated), which results in one or more Occurrences.

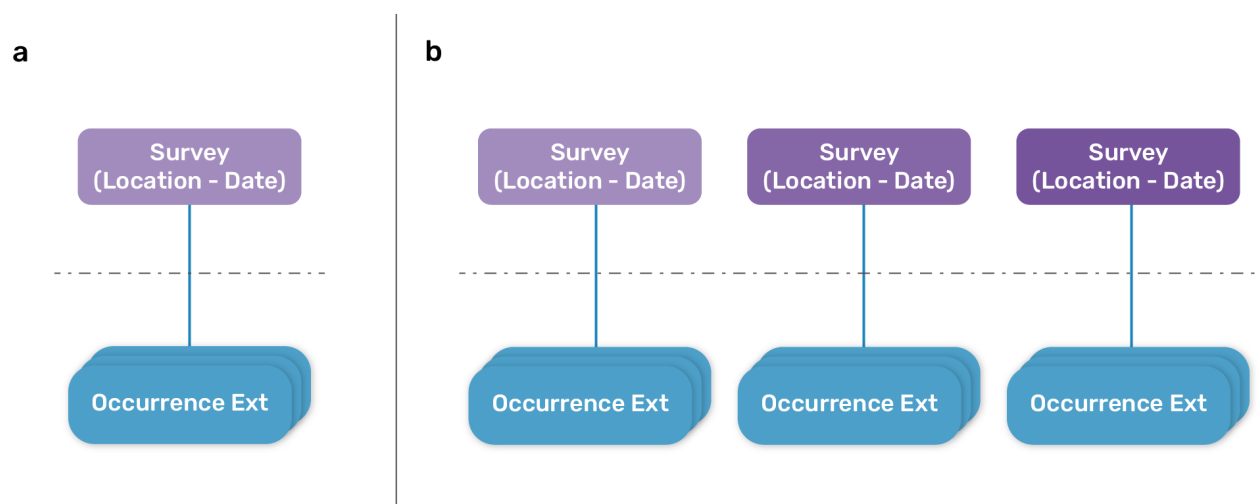


Figure 1. A simple schematic of a non-nested Event dataset (a) consisting of a single Event (purple box) with associated Occurrences related to the Event via the Occurrence extension (blue box) and (b) a series of individual Events (purple boxes) with associated Occurrences related to the appropriate Event via the Occurrence extension (blue boxes).

## 3.3. Nested datasets

More complex survey designs will require a nested structure. **Nested datasets** use parent-child relationships to capture information about more complex survey designs, such as datasets resulting

from repeated sampling Events and/or multiple sampling protocols. Creating nested Event levels may be important or even essential to relating the full story a dataset has to tell and to facilitating downstream analysis of the data by including the information necessary for connecting related records as part of the data.

### In a nested dataset:

- The top-most Event level does not have a parent Event but is parent to all Events beneath it.
- An Event may be parent to multiple child Events.
- All Events **except those at the lowest Event level** are considered the **parent Event** to any Event(s) beneath it.
  - **A parent Event must fully encompass its child Events spatially and temporally.** Specifically, the spatial extent and temporal interval of a parent Event must contain the spatial extents and temporal intervals of all of its children (see Section 3.2.1 Principle of spatiotemporal coverage in [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#)).
  - A **child Event** (an Event that is contained entirely within a single parent Event) may represent either multiple sampling sites, protocols, or repeated sampling at the same locality using the same protocol.
- Events **at the lowest hierarchical level** are never a parent Event.

**Each Event level should reflect a meaningful ecological or operational unit (e.g., spatial, temporal, or ecological) in the survey design.** An Event level should only be added if the addition of that Event level is necessary to facilitate data interpretation, downstream analysis, and/or linkage of information across data sources. **Do not create Event levels that are not necessary.**

Refer to [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#) (TDWG Humboldt Extension Task Group, 2024) for more information about creating nested data structures for Darwin Core datasets.

The goal in establishing a dataset structure is to keep it as simple as possible while still accurately representing the survey design. There may be multiple ways to structure a dataset and **there is no single correct dataset structure**. Further, identifying the data structure most appropriate for a dataset may not be a straightforward process. As a general guideline, dataset structure is most commonly defined as a function of sampling location, protocol, and date.

### 3.3.1. Simple nested data structures

Consider a hypothetical survey where two sampling protocols (Protocol a and Protocol b) are implemented at two different sites (Site 1 and Site 2). Both sites are sampled (site visits) twice (t1 and t2) using each of the protocols.

This survey dataset could be structured with two Event levels as shown in [Figure 2](#). Here, the highest Event level would consist of four Events representing each unique site-protocol combination: Site 1–Protocol a, Site 1–Protocol b, Site 2–Protocol a, Site 2–Protocol b. Events at the lowest Event level will represent site visits that occur on a particular date for each site-protocol combination. Organismal Occurrence information collected during each site visit is linked to the relevant site visit Event. This two Event level structure represents the simplest possible nested dataset structure with



only a single level of nesting.

It is ideal to structure a dataset such that each implemented protocol and unique site location is represented as a specific Event so that information from the same pool of species (i.e. location) and likelihood of detecting these species (i.e. protocol) is joined together by being part of the same Event. However, it is not always possible to disentangle information collected using multiple protocols.

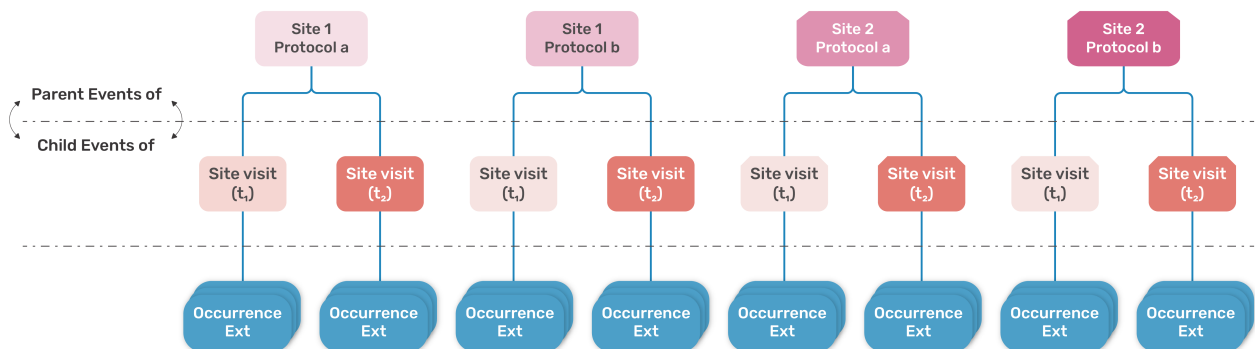


Figure 2. Simplified example schematic of a nested Event dataset consisting of a series of surveys conducted at two sites (Site 1 and Site 2) with two distinct sampling protocols (Protocol a, Protocol b) represented by the pink boxes. Surveys implementing each protocol are conducted at Sites 1 and 2 on two different dates (Site visit t1, Site visit t2; orange boxes). Associated Occurrences are related to the appropriate Event via the Occurrence extension (blue boxes).

### 3.3.2. Simple nested datasets with Project-level information

Surveys conducted as part of a larger or established network or project should report as much contextual information as possible to capture information about the project or network. Project-level information will always be shared at the highest Event level. This can be achieved in one of two ways:

- **By embedding project-level information within the highest existing survey Event level.** With the dataset presented in Figure 2, project-level information would be included with each of the four Site-Protocol Events.
- **By introducing a new parent Event level above all existing Events dedicated to capturing project-level information.** In the context of the example dataset presented in Figure 2, this would mean adding a third Event level to the dataset structure that is parent to all four Site-Protocol Events (see Figure 3). Creating a single parent Event is particularly useful option when a project will result in multiple, independent datasets. In this case, the Event identifier used for the project Event level can be used in all relevant datasets providing a means of identifying related datasets.

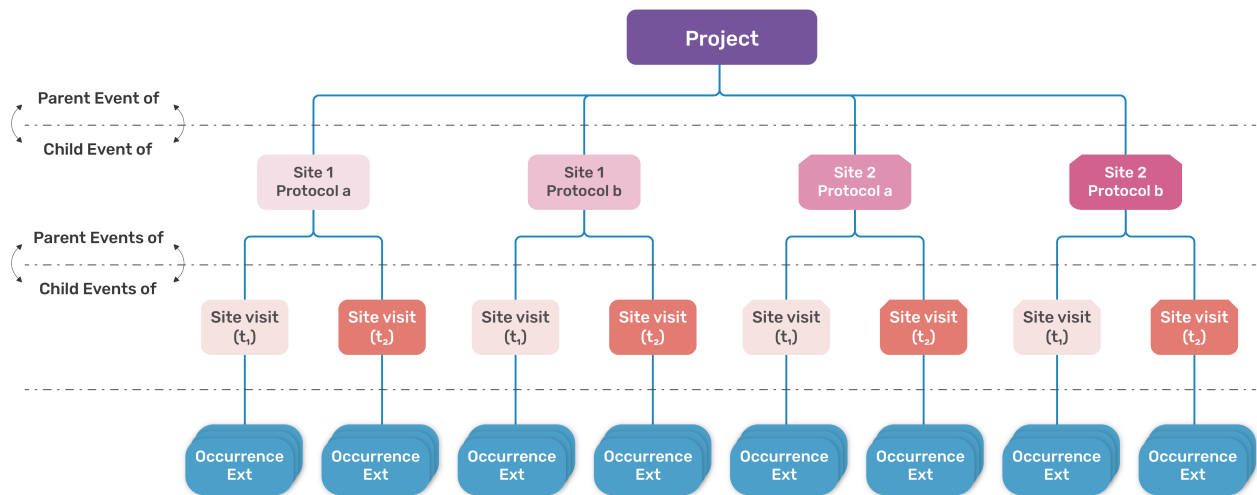


Figure 3. Simple nested hierarchy as presented in Figure 2 with the addition of a single Event that is parent to all other Events to consolidate all survey Events under the context of the broader project (purple box).

### 3.3.3. Deeply nested datasets

Although the recommendation is to keep dataset structure as simple as possible, more complex nesting may be necessary to accurately represent survey design and support data reuse. Added structural complexity can improve clarity when:

- multiple protocols are implemented within the same survey design,
- survey outputs include a mix of data types (e.g., specimen collections, field observations, observed co-occurrences),
- collected material contributes to downstream products (e.g., trait data, lab measurements, voucher specimens, media representations), or
- relationships among datasets need to be preserved or exposed (e.g., datasets resulting from different types of surveys within the same Project and/or at the same established survey sites).

For example, consider the dataset [Krill along the 110°E meridian: Oceanographic influences on assemblages in the eastern Indian Ocean, RV Investigator voyage IN2019\\_V03 \(2019\)](#), published by Ocean Biodiversity Information System (OBIS)-Australia. The dataset contains information about a zooplankton survey conducted by the [CSIRO Marine National Facility](#) in the eastern Indian Ocean in 2019. The survey consisted of daytime and nighttime sampling at 20 locations (stations) along an established transect. As illustrated in [Figure 4](#), this dataset could be structured as a non-nested dataset (Figure 4a) or as nested dataset (Figures 4b-d); and, as a nested dataset, the structure could be simple (Figures 4b and c) or more deeply nested with more than two Event levels (Figure 4d).

- **Non-nested dataset structure** (Figure 4a): As a non-nested dataset, each sampling at a given station at a particular date and time would be a unique Event with no obvious link to other Events in the dataset beyond being part of the same dataset. Implementing this structure is the simplest approach to sharing data from the survey, however, without any nesting of Events, it may be difficult for data users to understand the relationships between survey Events. Associated Occurrences are related to the appropriate Event via the Occurrence extension.
- **Simple nested dataset structure** (Figure 4b): An alternative a simple nested dataset structure could consist of two Event levels. The highest Event level would capture information about the survey stations, where each of the 20 survey stations would be a unique, unrelated parent Event to the relevant daytime and nighttime sampling Events. Associated Occurrences would be related to the appropriate Event via the Occurrence extension.

- **Simple nested dataset structure** (Figure 4c): As a simple nested dataset, the data structure would consist of two Event levels with the highest Event level capturing information about the overall cruise or campaign and second Event level represents the daytime and nighttime sampling events at each station as a series of unique Events. Associated Occurrences are related to the appropriate Event via the Occurrence extension.
- **Deeply nested dataset structure** (Figure 4d): As a more deeply nested dataset, the structure would consist of three Event levels: the highest Event level represents the Survey (that is, the overall cruise or campaign); the middle Event level represents each of the 20 survey stations; and, the lowest Event level represents the daytime and nighttime sampling Events at each station. Note that the child Events of each parent Event are used to report independent replicates of the same type within the same parent Event and/or to preserve individual sampling units. Associated Occurrences are related to the appropriate Event via the Occurrence extension.

If the survey itself was a unique Event, the simpler two Event level structure (e.g., Figures 4b and 4c) would likely suffice. However, the stations sampled during the survey are standard sampling locations used in other survey efforts not covered by this dataset. To make it easier to link information from this dataset to data from other surveys conducted at the same localities, a more complex nested structure was chosen by the data publisher.

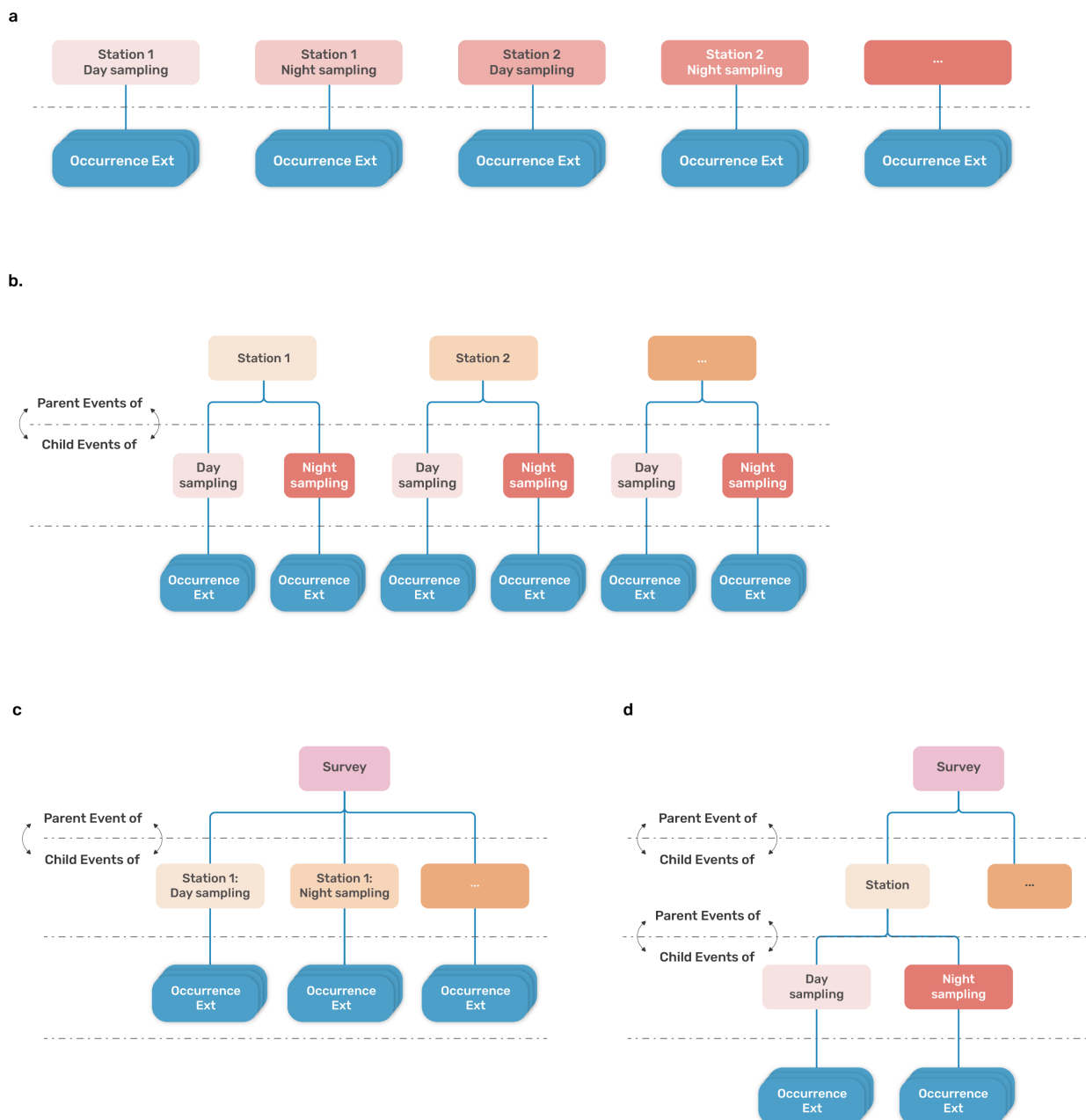


Figure 4. Four potential dataset structures for a zooplankton survey conducted by CSIRO at 20 stations, each sampled once during the day and once at night: non-nested structure (a), simple nested structure (b and c), and complex or deeply nested structure (d).

### 3.3.4. Constructing a dataset schematic

As noted in the previous section, some datasets may be very simple and have no hierarchical structure (**non-nested datasets**) with singular observations of individual taxa at a single location. Others may be complex and hierarchically structured (**nested datasets**), with a series of nested survey Events (e.g., sampling designs with traps within plots within sites). Multiple structural scenarios may fit a dataset, particularly for more complex data resulting from ongoing monitoring or repeated sampling efforts. We recommend keeping the structure as simple as possible. Refer to [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#) for additional guidance on how to capture the details of nested observations (dwc:Event hierarchies).

Creating a schematic of the dataset hierarchical structure such as in Figures 1-4 is particularly useful in exploring and effectively capturing the survey design that generated the data collected. Once the dataset structure is identified, the schematic can be expanded to identify which extensions (e.g.,

Humboldt, Occurrence, extended measurement or fact) are needed, if any, and where they will link (see Box 1 below and Figure 1 of [De Pooter et al. 2017]). After, you can proceed with mapping your data to the [DwC Event Core](#) and the [Humboldt extension](#) as described in the following sections.

### Box 1. National Science Foundation's National Ecological Observatory Network (NEON) example

The structure of an example nested dataset from the U.S. National Science Foundation's National Ecological Observatory Network (NEON), a long-term ecological data collection facility, is presented in [Figure 5](#). This structure describes tick-pathogen data derived from two interconnected NEON datasets: [Ticks sampled using drag cloths \[NEON 2025\]](#) and [Tick pathogen status \[NEON 2025\]](#). Adapting the survey design of these two datasets to a deeply nested structure allows preservation of the associations between pathogen detections and their corresponding host ticks across collection areas which would otherwise be separated across two non-nested datasets.

The schematic in [Figure 5](#) is a general interpretation of the dataset structure and follows NEON's standard survey design [Thorpe et al. 2016]. The NEON system is broadly divided into 20 ecoclimatic domains across the United States of America and Puerto Rico. Across these domains, NEON has established a total of 81 field sites (47 terrestrial and 34 aquatic), which serve as representative sampling locations within each domain. Within each field site, spatial sampling units, such as plots, segments, or reaches, are established based on the site type and the requirements of individual protocols. To simplify standardization of information across datasets, NEON's biological datasets are structured such that locality information (domain, site, sampling unit) is contained in the highest three event levels and information specific to individual site visits is reported at the lowest Event levels. At each sampling Event, information about the sampling context is reported using [DwC Event core](#) and [Humboldt extension](#) terms. Information that cannot be reported using those terms is reported using the [extended measurement or fact \(emof\) extension](#). This is the information that is least likely to change through time and would make it easier to aggregate information across their own datasets. Depending on data collected, other extensions can also be added, such as the [simple multimedia](#) or [DNA-derived data](#) extensions. Associated Occurrences are related to the appropriate Event via the Occurrence extension.

Note here that reporting of Occurrence information is illustrated using two instances of the [Occurrence extension](#), one for ticks and another for pathogens. While Occurrence information is most commonly shared using a single table, it can be shared using multiple tables. For the purposes of preparing this dataset as a case study for the guide, tick and pathogen Occurrence data were kept separate to more clearly illustrate that pathogen records are derived from samples collected from the ticks. This relationship is communicated using the [resource relationship extension](#).

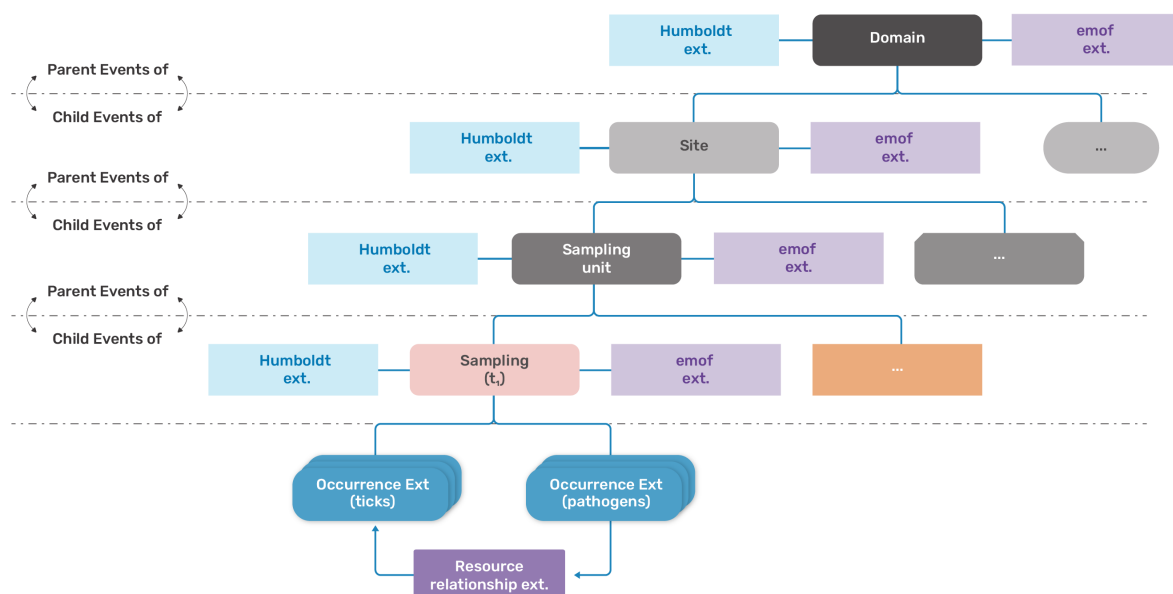


Figure 5. Generalized dataset structure for NEON biological datasets. The structure is a deeply nested dataset to more precisely capture details about NEON's standardized survey design. The locality component encompasses 3 hierarchical levels to highlight the shifts in spatial scope from ecoclimatic domains to sites to individual sampling units (gray boxes) with individual sampling Events or surveys conducted at a sampling unit (orange boxes) representing the lowest Event level. Additional information about each location or survey Event is shared using the Humboldt extension (light blue boxes) and the extended measurement or fact extension ('emof ext.', light purple boxes). Occurrence information associated with a sampling Event is reported using two separate tables, one each for tick occurrence information and the second for pathogens identified in tick samples (dark blue boxes), to simplify data reporting and to illustrate implementation of the resource relationship extension (purple box) used to link pathogens with the tick samples from which they were identified.

## 4. Resource metadata



Resource metadata information should be saved to the DwC-A dataset metadata file (eml.xml).

**Resource metadata** provides project- and/or dataset-level information for potential data users to understand the context of a dataset. GBIF's metadata schema is based [Ecological Metadata Language \(EML\)](#), a metadata standard administered and maintained by [The Knowledge Network for Biocomplexity](#), which captures information about an ecological dataset in a series of modular and extensible XML documents. Each Darwin Core Archive must include a resource metadata written in XML format: [eml.xml](#).

GBIF currently requires 8 dataset-level metadata terms (see [Data Quality Requirements for Sampling Events](#) for more information):

- **title:** This is the title under which the dataset will be published at [gbif.org](#). The title should be brief, but long enough and descriptive enough to characterize the dataset in an international context and distinguish it from similar datasets from other institutions.
- **description:** A brief, textual description of the dataset. This may include an extended version of the title, a description of the geographic, temporal and taxonomic scope(s) of the dataset, information about the methodology implemented and purpose of the underlying data compilation (e.g. protected habitat surveillance, faunistic inventory, deep sea trawl data, survey steps or gear used), relevant literature references, and any other information you consider relevant to characterize the dataset. This is, in essence, a resource abstract.
- **publishing organization:** The name of the institution or organization that will be listed as the data publisher at [gbif.org](#). The publishing organization is the institution which holds or owns the dataset and is in charge of its contents and maintenance.
- **type:** Type refers to the dataset structure reflecting the level of detail captured in the dataset. In GBIF, four types of datasets are currently accepted: sampling event, occurrence, checklist, and metadata. Type for survey and monitoring datasets is [samplingEvent](#).
- **license:** A machine-readable statement of the rights and intended use attached to the published dataset. GBIF supports the following Creative Commons categories: CC0, CC BY, and CC BY-NC (see [GBIF Terms of use](#)).
- **contact(s):** The contact field contains contact information for a dataset. This is the person or institution to reach out to with questions about the use or interpretation of a dataset. The information for at least one contact is required to ensure the possibility of communication about the dataset. Minimum required information for resource each contact is name and email address.
- **creator(s):** A resource creator is the person(s) or organization(s) responsible for creating a resource. Contact information for at least one dataset creator is required. The minimum required information for each dataset creator includes name and email address for at least one contact.
- **metadata provider(s):** The metadata provider is the person or organization responsible for providing documentation for a resource. At least one metadata provider must be listed. The minimum required information for each metadata provider is name and email address.

These 8 terms must be populated in order to successfully publish a dataset to GBIF. See the [GBIF Metadata Profile – How-to Guide](#) for comprehensive guidelines and a list of all available resource metadata terms [\[GBIF 2011\]](#).



## 5. Survey Event data: capturing the context of biological surveys and monitoring data



The contextual information about survey Events should be saved to the DwC-A **event** table.

This section will guide you through the process of mapping Event-level data specifically related to survey structure, location, protocols, scopes, and effort.

### *About DwC terms in this document*

Each term in this document is linked with its respective term **internationalized resource identifier (IRI)** alias (ex., `eco:protocolNames`). Always use these links to refer to the definition, comments, and examples provided when populating a term.



The terms to be used to describe Event-level information are a combination of Darwin Core Event class and Humboldt Extension terms:

- A comprehensive list of Darwin Core Event class terms is available on the **Darwin Core Event** page in the GBIF Registry of Schemas.
- The full suite of Humboldt extension terms is available on the **Humboldt Ecological Inventories** page in the GBIF Registry of Schemas and the **Humboldt Extension Quick Reference Guide**.

### Data mapping tips

Survey Event data should be saved to the DwC-A **event** table. This includes DwC Event terms (any term preceded by **dwc**, e.g., `dwc:eventID`) and Humboldt extension terms (any term preceded by **eco**, e.g., `eco:protocolNames`).

Populate all terms for which information is available.

**Paired terms** must be populated together. These terms are designed to offer data publishers some level of flexibility in reporting data. Paired terms are most common in terms available for reporting a variable value and associated unit of measure (for example, `dwc:sampleSizeValue` and `dwc:sampleSizeUnit`).

#### No data, missing data, and data values of 0

- Cells with a value of 0 (zero) should be explicitly populated as 0.
- Cells with missing data or **NULL** values should be left empty.
- Terms for which there is **no data** to share at any hierarchical level can be excluded from the data table.

#### Populating terms across Event levels (e.g., from parent Event to child Event)

- Each Event can have its own set of attributes and measurements which can be captured using the Humboldt and/or other extension(s) and be unambiguously linked to the corresponding Event through the appropriate `dwc:eventIDs`.
- **Terms should contain data clearly (explicitly) reported at every Event level in the hierarchy to which they directly apply.** This means that when publishing a data export,

- a child Event must **not** be assumed to inherit the value of any property of any of its parent Event(s). If a value is in common between a child Event and its parent Event, the value should be reported explicitly in both Event records (see Section 3.2.2 Principle of applicability in [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#)).
- a parent Event should **not** be populated by deriving or summarizing information from child Events; rather, fields should be explicitly populated with information directly appropriate to the nature and level of the Event itself (see Section 3.2.3 Principle of non-derivation in [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#)).

Refer to [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#) for more guidance in populating Humboldt extension terms across Event levels.

## 5.1. Survey design

**Survey design** is the strategy underpinning a biological survey. It details the sampling method implemented in a particular survey including how any stations, plots, traps, sensors, and/or transects are positioned. Historically, only 2 terms were available to structure and relate different levels of survey design in a dataset: `dwc:eventID` and `dwc:parentEventID`. One additional Darwin Core Event term, `dwc:fieldNumber`, provided a means by which to relate a sampling Event with a dataset- or project-specific field number. The Humboldt extension provides an additional 2 terms (`eco:siteCount` and `eco:siteNestingDescription`) to better support complex or nested survey designs.



### Event data in GBIF

- Any dataset to be published using the DwC Event core must have at least one Event record.
- Each `dwc:eventID` in a dataset must be unique within the dataset. Use of a persistent globally unique identifier (GUID) is recommended to ensure that the GUID is unique across all datasets. A unique `dwc:eventID` should be reused between datasets where appropriate (for example, where data collected during the same sampling event are published as multiple datasets). See [A Beginner's Guide to Persistent Identifiers](#) for guidance in creating persistent identifiers. Note that your field numbers should be reported using `dwc:fieldNumber`.
- An Event is not required to have associated organism Occurrence data. If organism Occurrence or non-detection data are available, they will be linked via the `dwc:eventID` in the **occurrence** table using the **occurrence extension**.
- Other DwC Event extensions, including **occurrence**, **extended measurement or fact**, and **relevé** extensions, can be linked to any appropriate Event via the `dwc:eventID`.

### Non-nested datasets

- Each Event in a non-nested dataset must be assigned a unique `dwc:eventID`.
- Non-nested datasets will not have a `dwc:parentEventID`.

### Nested datasets

Nested hierarchies are established by relating a child Event to a parent Event through the child Event's `dwc:parentEventID`. As such, these more complex datasets require use of both `dwc:eventID`

and `dwc:parentEventID`.

- Each Event in a nested dataset must have a unique `dwc:eventID`.
- Each child Event should include the `dwc:parentEventID` of its parent in `dwc:parentEventID`.

In practice, this means that the parent and the child will each have a unique `dwc:eventID`. To create the parent-child relationship, the parent Event's `dwc:eventID` will be also be reported as the child Event's `dwc:parentEventID`.

*Simple example illustrating how a parent-child relationship between two Events would look using Event identifiers.*

<code>dwc:parentEventID</code>	<code>dwc:eventID</code>
	<code>survey2022</code>
<code>survey2022</code>	<code>survey2022_a-2</code>

In addition to Event and parent Event identifiers:

- **Site count and site nesting description:** Nested datasets should include the total number of sites sampled in `eco:siteCount` and provide a textual description of the survey design or site sampling structure using `eco:siteNestingDescription` for each parent Event for which the information is available.
- **Field number:** If the survey data include a field number for a specific Event, this should be shared using `dwc:fieldNumber`.

*Event hierarchy terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Required	<code>dwc:eventID</code>	<code>survey2022_a-2</code>
Required for nested datasets	<code>dwc:parentEventID</code>	<code>survey2022</code>
Recommended	<code>eco:siteCount</code>	75
	<code>eco:siteNestingDescription</code>	25 survey sites each with 3 1m <sup>2</sup> quadrats
Share if available	<code>dwc:fieldNumber</code>	RV Sol 87-03-08

## 5.2. Project information

If the survey(s) being reported were part of a larger Project, four terms are available to capture the project name(s) and funding institution(s).

- **Project title:** The official name(s) of the project(s) that contributed to the creation of the dataset should be shared as a concatenated list with values separated using a pipe separator `|` in `dwc:projectTitle`.
- **Project ID:** A list, concatenated and separated using a pipe separator `|`, of the globally unique identifiers for the project(s) that contributed to the creation of the dataset should be reported in `dwc:projectID`.
- **Funding attribution:** The official name(s) of the funding body or bodies that provided funding for the survey(s) resulting in the creation of the dataset should be shared as a concatenated list with values separated using a pipe separator `|` in `dwc:fundingAttribution`.
- **Funding attribution ID:** A list, concatenated and separated using a pipe separator `|`, of the globally

unique identifiers for the funding organizations or agencies that supported the project can be provided in `dwc:fundingAttributionID`.

*Project terms, their recommended usage (status), and example data entries*

Status	Term	Example entry
Share if available	<code>dwc:projectTitle</code>	Scalidophora i Noreg, Biowide
	<code>dwc:projectID</code>	RCN276730   Artsproject_7-24, <a href="https://arvenetternansen.com/">https://arvenetternansen.com/</a>
	<code>dwc:fundingAttribution</code>	Norges forskningsråd
	<code>dwc:fundingAttributionID</code>	<a href="https://ror.org/00epmv149">https://ror.org/00epmv149</a>   <a href="https://ror.org/04jnzhb65">https://ror.org/04jnzhb65</a>

## 5.3. Survey site

An Event **site** is the location at which observations are made or samples and/or measurements are taken. Sharing thorough information about a sampling Event site, including description, locality, and vegetative cover provides critical context to potential data users about conditions in which a survey was conducted. Information about the location of each survey site such best-practice georeferences, site description (locality name, habitat type, microhabitat), and environmental data (e.g., physical parameters, vegetarian, water quality) should be populated at for each Event for which the information is available.



The Darwin Core site terms listed in this section are not comprehensive. Explore all [Darwin Core Location class terms](#) and the [Humboldt Extension site terms](#).

### 5.3.1. Site description

Additional context about a survey site can be reported through myriad terms for every Event that the information is available, including:

- **Site names:** survey site names can be reported using `eco:verbatimSiteNames`. A concatenated list of site names can be provided at higher Event levels with values separated using a pipe separator, `|`.
- **Habitat:** reported habitat at a survey site should be recorded in `dwc:habitat`. A concatenated list of habitats can be provided at higher Event levels with values separated using a pipe separator, `|`. Use of a controlled vocabulary is recommended. Note that a single controlled vocabulary does not exist for this term yet but attempts to classify habitat have been and continue to be made (for example, see [\[Keith et al. 2020\]](#) or [\[Campbell et al. 2021\]](#)).
- **Weather:** reported weather during a survey Event should be reported using `eco:reportedWeather`. If you have detailed weather data (e.g., weather station or data logger produced data) archived elsewhere, you may provide a link here.
- **Extreme conditions:** reported extreme conditions at a site at the time of the survey should be recorded in `eco:reportedExtremeConditions`.
- **Verbatim site description:** verbatim comments (e.g., the original textual description) about a site or sites should be recorded in `eco:verbatimSiteDescriptions`.

These terms should be populated for each individual Event for which the information is accurate.

*General event site terms, their recommended usage (status), and example data entries*

Status	Term	Example entry
Share if available	eco:verbatimSiteNames	Trap_18 Trap_27 Trap_54 Trap_96, Annala   Kumpula
	dwc:habitat	Ephemeral wetland
	eco:reportedWeather	{"minimumTemperatureInDegreesFahrenheit": 18, "maximumTemperatureInDegreesFahrenheit": 32}
	eco:reportedExtremeConditions	Site flooded
	eco:verbatimSiteDescriptions	Coastal sand dunes at dry oak forest edge. Vegetation: <i>Ammophila arenaria</i> , <i>Betula pendula</i> , <i>Leymus arenarius</i> , <i>Pinus sylvestris</i>

### 5.3.2. Site locality

The geographic location and extent of each survey site should be reported. Five terms are currently recommended for Event datasets:

- **Location ID:** a unique identifier for each survey site should be shared in `dwc:locationID`. If a site is visited repeated (as in long-term monitoring and other repeated survey efforts), `dwc:locationID` should be consistent across Events within a dataset and across datasets in situations where the same survey sites are visited in other datasets.
- **Country code:** the ISO two letter code for the country, region, or economy in which a survey takes place should be provided in `dwc:countryCode`.
- **Latitude-longitude:** The decimal latitude and longitude and geodetic datum location of each survey site should be reported in `dwc:decimalLatitude`, `dwc:decimalLongitude`, and `dwc:geodeticDatum`. All three terms should be populated together.
  - If the geographic coordinates of your dataset are not in decimal latitude and decimal longitude format, use the terms `dwc:verbatimLatitude`, `dwc:verbatimLongitude`, and `dwc:verbatimCoordinateSystem` to report geographic location instead.
  - Note that this is a minimum recommendation and does not make data fit for the maximum number of purposes. It is highly recommended to provide georeference information that follow best practices.

### 5.3.3. Survey site area

Reporting additional information about the areas targeted for sampling and the area(s) actually sampled during a survey is recommended to provide greater context about the geospatial scope of a survey. The Humboldt extension includes two sets of paired terms to report the survey area of an Event: *geospatial scope* terms and *total area sampled* terms.

- **Geospatial scope** terms (`eco:geospatialScopeAreaValue` and `eco:geospatialScopeAreaUnit`) define the geospatial scope or extent of a survey or sampling Event. Geospatial scope terms can be applied at any Event level and should report the entire area considered for the survey.
- **Total area sampled** terms (`eco:totalAreaSampledValue` and `eco:totalAreaSampledUnit`) report the area actually sampled during an Event. Total area sampled terms can be populated at any Event level but are most commonly applied at lower Event levels to, for example, capture the survey extent of a single plot or (at higher Event levels) the cumulative area surveyed in a series

of plots within a site.

In **non-nested event datasets**, geospatial scope terms and total area sampled terms may contain the same values.

In **nested datasets**, geospatial scope terms will be equal to or greater than the area values shared in total area sampled terms. See Box 2 for an example.

**If the surveyed unit is not an area** (e.g.,  $\text{km}^2$  or  $\text{m}^2$ ), `dwc:sampleSizeValue` and `dwc:sampleSizeUnit` should be used instead. Examples include:

- point locations (such as a sensor or trap),
- distances (such as transect lengths), and
- volumetric measures (such as a filtered volume of water in a zooplankton haul).

### Box 2. Biowide project example

Consider the **Biowide project** which surveyed 130 40×40m plots across Denmark.

Here, the project-level parent Event would report the full geographic extent of Denmark: `eco:geospatialScopeAreaValue` = 42934 and `eco:geospatialScopeAreaUnit` =  $\text{km}^2$  and the sum of sampled areas: `eco:totalAreaSampledValue` = 208000 and `eco:totalAreaSampledUnit` =  $\text{m}^2$ .

For each of the 130 associated child Events representing the individual plots, `eco:geospatialScopeAreaValue` and `eco:geospatialScopeAreaUnit` would be left empty because `geospatialScope` was a characteristic of the higher-level survey design not the individual survey site visits. For the plots, the area of the site would be `eco:totalAreaSampledValue` = 1600 and `eco:totalAreaSampledUnit` =  $\text{m}^2$ .

### 5.3.4. Additional survey site information

- **Survey site geometry:** If available, the geometry of a survey site area should be shared using `dwc:footprintWKT` and `dwc:footprintSRS`.
- **Verbatim site location information:** A more general text description of the site location, if available, can be shared using `dwc:locality`.

*Event site geographic locality and scope terms and their recommended usage (status), namespace abbreviation, and example data entries.*

Status	Term	Example entry
Recommended	<b>dwc:locationID</b>	Trap_138
	dwc:countryCode	SE
	dwc:decimalLatitude	59.3168
	dwc:decimalLongitude	18.0627
	dwc:geodeticDatum	epsg:4326
	eco:geospatialScopeAreaValue	580000
	eco:geospatialScopeAreaUnit	km <sup>2</sup>
	eco:totalAreaSampledValue	1600
	eco:totalAreaSampledUnit	m <sup>2</sup>
	dwc:sampleSizeValue	200
	dwc:sampleSizeUnit	m <sup>3</sup>
Share if available	dwc:footprintWKT	POLYGON ((10 20, 11 20, 11 21, 10 21, 10 20))
	dwc:footprintSRS	epsg:4326
	dwc:locality	Agriculture site, Kongskilde Friluftsgård, Zealand

### 5.3.5. Vegetation cover

If vegetation cover data are available for a site (for example, if a relevé was conducted or if a textual site description was provided), it can be reported in three ways:

- **Verbatim vegetation cover:** a verbatim or textual description of vegetation cover can be captured using `eco:verbatimSiteDescriptions`.
- **Percent vegetation cover:** simple percent vegetation cover can be recorded as structured data using the [extended measurement or fact extension](#). Data reported using `emof` should be linked to the appropriate Event using `dwc:eventID`. See the ['Extended measurement or fact \(eMoF\) extension'](#) section for details on using the extension.
- **Vegetation plot survey:** vegetation plot survey information (that is, data collected during a relevé) can be reported using the [relevé extension](#). Data from individual relevé's should be linked to the appropriate Event using `dwc:eventID`. See the ['Relevé extension'](#) section for details on using the extension.

There is no single best method of reporting vegetation cover information for a site, although it is recommended to choose the most explicit method possible based on the type of information available.

If vegetation cover is reported using one of the three methods described above, then `eco:isVegetationCoverReported = true`; otherwise, `eco:isVegetationCoverReported = false`.

## 5.4. Survey date and time

Complete and accurate reporting of the temporal scope of a survey is crucial to asserting Event structure and providing key contextual information about sampling conditions.

Each Event should include a date or date range in `dwc:eventDate`. **Nested datasets** should, at the parent Event level, report a date range encompassing the dates of all relevant child Events.



The time and duration of each Event should be reported using `dwc:eventTime` and the paired terms `eco:eventDurationValue` and `eco:eventDurationUnit` respectively.

Refer to GBIF's technical documentation on [date and time interpretation](#) for more guidance on reporting Event dates and times.

*Event date and temporal scope terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Required	<code>dwc:eventDate</code>	2018-08-29
Recommended	<code>dwc:eventTime</code>	08:00Z
	<code>eco:eventDurationValue</code>	1
	<code>eco:eventDurationUnit</code>	hour

## 5.5. Methodology or sampling protocol

**Sampling protocols** provide the details of how a survey was conducted. Protocol information should be a detailed, step-wise description outlining all the details about the data collection process necessary to ensure repeatability of the implemented methodology. Clear communication of a sampling protocol or the method(s) implemented during a survey or monitoring effort guarantees consistency, accuracy, and reliability in the data collected. This information further ensures reproducibility and reusability of a dataset, and facilitates data aggregation, integration, and subsequent analysis.



Sampling protocol terms should be populated for every Event regardless of hierarchical level as inheritance in either direction should not be assumed or inferred between Event levels.

### 5.5.1. Event type

Biological survey Event data can result from a wide variety of effort types (e.g., Bioblitzes, inventories, monitoring schemas, expeditions). The nature of the survey Event should be reported using `dwc:eventType`.

#### Identifying Event type

`dwc:eventType` should provide a high level overview or broadly categorize the type of survey without being so specific as to overlap with sampling protocol. There is no single, standardized vocabulary for `dwc:eventType`. If your organization or community has a controlled vocabulary, it is recommended to apply terms from that. Otherwise, you can refer to the common Event types below for guidance. More than one term may apply to an Event; choose the term that fits most closely.

- **Project:** Projects are structured initiatives with an explicitly stated objective or suite of objectives and with clear targets, timelines, and deliverables. Projects typically are linked to non-biological information identifying participating organizations and people (agents), funding agencies, and other high-level administrative information. Biological sampling may be only one facet of a project's scope. **Project** as a `dwc:eventType` is typically most appropriate only at the highest Event level in a nested dataset.
- **Expedition:** An expedition is an organized information gathering venture that inherently includes multiple sampling Events and event types. Expeditions may include multiple

taxonomic and/or organismal scopes, any number of documented sampling protocols, and varying degrees of complexity in survey design. **Expedition** as an `dwc:eventType` is typically most appropriate at higher Event levels in nested hierarchies.

- **Survey:** A survey is a broad but systematic effort to collect information about the biological organisms in a specific area at a given time. Surveys typically included at least one documented protocol and may or may not have an explicitly defined taxonomic and/or organismal scope. **Survey** is the most general Event type term and can be applied as an `dwc:eventType` at any Event level.
- **Inventory:** An inventory is a comprehensive, focused survey of the taxa present in a specific area over an explicit period of time. Inventories typically have an explicit taxonomic and/or organismal scope and a well-defined protocol. **Inventory** is typically most appropriate as an `dwc:eventType` at lower Event levels in nested hierarchies.
- **Bioblitz:** A bioblitz is a survey Event aimed at finding and identifying as many species as possible in a specific area over a (typically) short, contiguous period of time. Bioblitzes often include participants (agents) with a wide range of backgrounds and levels of expertise in biodiversity sciences including formal biologists as well as the broader, general public. **Bioblitz** as an `dwc:eventType` is typically most appropriate at lower Event levels in nested hierarchies.
- **Site visit:** A site visit is a single survey, inventory, or sampling at a pre-established geographic location at a discrete time. **Site visit** as an `dwc:eventType` is typically most appropriate at the lowest event level in a nested hierarchy.
- **Sample:** A survey event denoted by the specific act of collecting physical samples resulting in material specimen. A sampling `dwc:eventType` is a specific implementation of a survey Event. 'Sample' as an `dwc:eventType` is typically most appropriate at lower (child) Event levels in a nested hierarchy.
- **Sensor:** The detection of an Occurrence (or a group of related occurrences such as a time series or group of organisms) by means of a sensor. A sensor may be static (e.g., camera traps) or mobile (e.g., drones) external to an organism, or it may be attached to an organism (e.g. radio collar). 'Sensor' as an `dwc:eventType` is typically most appropriate at lower (child) Event levels in a nested hierarchy.

## Inventory Event types

If `dwc:eventType` = **inventory**, the type(s) of search implemented (e.g., restricted search, open search, opportunistic search, trap or sample, compilation) must be reported in `eco:inventoryTypes`.

If `eco:inventoryTypes` = **compilation**, the compilation type should be reported using `eco:compilationTypes` and data sources listed in `eco:compilationSourceTypes`.

- A is a summary inventory resulting from the combination of multiple existing inventories (as described in [Guralnick et al 2018]). Compilations are aggregates of multiple studies and may combine surveys employing different protocols, processes, and observers, often with variable reporting of the methods employed or other compiled data sources and literature searches.

*Event type terms, their recommended usage (status), and example data entries*

Status	Term	Example entry
Recommended	<code>dwc:eventType</code>	<b>Inventory</b>

Status	Term	Example entry
Recommended if applicable	eco:inventoryTypes	Open search
	eco:compilationTypes	compilationOfExistingSourcesAndSamplingEvents
	eco:compilationSourceTypes	museumSpecimens   literature

### 5.5.2. Sampling protocol

Four protocol terms exist; however, only 1 term is currently required to publish an Event dataset in GBIF: `dwc:samplingProtocol`. This requirement is because the initial Darwin Core Event classification only included the one term. The Humboldt extension introduced an additional three terms to capture information about sampling protocol in a more explicit manner:

- `eco:protocolNames`
- `eco:protocolDescriptions`
- `eco:protocolReferences`

#### Updated recommended best practice for reporting sampling protocol

- Report the name(s) of the sampling protocol in `eco:protocolNames`.
- Capture details of the methods or protocols implemented in `eco:protocolDescriptions`.
- Provide citation information for the protocol in `eco:protocolReferences`.
- Retain `dwc:samplingProtocol` as a verbatim field of `eco:protocolNames`.

*Survey Event protocol terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Required	<code>dwc:samplingProtocol</code>	Visual survey

Status	Term	Example entry
Recommended	eco:protocolNames	Visual survey
	eco:protocolDescriptions	For each site a total list of lichen species (lichenized fungi) was produced based on a careful examination of soil, wood, stone surfaces and bark of trees up to 2m at three time periods: October-November 2014, February-December 2015 and March and May 2016. Specimens that were not possible to identify with certainty in the field were sampled and subsequently identified in the laboratory. For each species the substrate, e.g. phorophyte (host) species was recorded. All records were registered in <a href="http://www.svampeatlas.dk">www.svampeatlas.dk</a> , and the nomenclature used is in accordance with this database.
	eco:protocolReferences	See Appendix B of Brunbjerg, A.K., Bruun, H.H., Brøndum, L. et al. A systematic survey of regional multi-taxon biodiversity: evaluating strategies and coverage. BMC Ecol 19, 43 (2019). <a href="https://doi.org/10.1186/s12898-019-0260-x">https://doi.org/10.1186/s12898-019-0260-x</a>   <a href="https://doi.org/10.17504/protocols.io.kxygx3jwkg8j/v1">https://doi.org/10.17504/protocols.io.kxygx3jwkg8j/v1</a>

### 5.5.3. Absences

**Organismal absences** are defined here as the lack of detection of organisms that are members of an explicitly stated target taxonomic scope. Absence information is critical to understanding species' biogeography, modeling species' responses to climate- and human-induced environmental change, conservation planning and resource management, monitoring and restoration efforts, eradications or reintroductions, and other aspects of biodiversity dynamics.

- If the dataset includes absence information for one or more organisms (to be reported in the **occurrence** table as `dwc:occurrenceStatus = absent`), then `eco:isAbsenceReported = true`.
- A list of absent taxa can be provided using `eco:absentTaxa` for all relevant Events. Best practice is to use scientific names to report absent taxa.
  - Absences should only be reported for taxa within the stated taxonomic and/or organismal scope of a survey and should use scientific nomenclature.
  - Absence cannot be asserted for bycatch.

See the section '**Reporting absences**' for details on reporting absence information at the Occurrence level.

### 5.5.4. Abundance

**Abundance** is a quantitative measure of the same taxonomic designation in a particular area at a specific time. Abundance data are a key indicator of ecological health. They are necessary for evaluating ecological patterns and dynamics, managing invasive species, informing effective habitat and ecosystem management, and for practical tasks such as quantifying existing resource.

- If the dataset includes any abundance information, `eco:isAbundanceReported` = **true** for all appropriate Events. If there is an abundance cap (that is, if there was a designated maximum value at which abundance was no longer counted), then `eco:isAbundanceCapReported` = **true** and the value of the cap should be reported in `eco:abundanceCap`.
- If there is no abundance cap, then `eco:isAbundanceCapReported` = **false**.

See the section '**Abundance information**' for details on reporting absence information at the Occurrence level.

*Absence and abundance terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Recommended	<code>eco:isAbsenceReported</code>	<b>true or false</b>
	<code>eco:isAbundanceReported</code>	<b>true or false</b>
	<code>eco:isAbundanceCapReported</code>	<b>true or false</b>
Share if available	<code>eco:absentTaxa</code>	
	<code>eco:abundanceCap</code>	<b>5</b>

### 5.5.5. Material samples

A **material sample** is a physical entity ‘... that represents an entity of interest in whole or in part’ (see `dwc:MaterialSample`). Essentially, material samples are specimens collected during a survey. A material sample may consist of an entire organism, part of an organism, or a genetic sample, or even multiple organisms not necessarily of the same taxonomic designation.

If the dataset includes at least one specimen from which a material sample was taken, for each relevant Event:

- `eco:hasMaterialSamples` = **true** and
- the type(s) of materials collected should be listed in `eco:materialSampleTypes`.

If the dataset or Event does *not* include material samples, `eco:hasMaterialSamples` = **false**.

### 5.5.6. Vouchers

A **voucher** is a physical specimen or material sample collected and accessioned into a museum collection in support of a specific project or survey.

If the dataset has vouchers, for each relevant Event:

- `eco:hasVouchers` = **true**, and
- a list of institutions housing them should be shared in `eco:voucherInstitutions`.

If the dataset or sampling event does *not* include vouchers, `eco:hasVouchers` = **false**.

### 5.5.7. Sensitive data: data generalization & information withheld

Although the general recommendation is to share all biodiversity data available at its highest spatio-temporal resolution, situations exist where it is necessary to generalize data prior to sharing a dataset publicly or even withhold information completely. Two terms are available to communicate if data are generalized or withheld in a dataset: `dwc:dataGeneralizations` and `dwc:informationWithheld`.

While it is the responsibility of the publisher to protect sensitive species occurrence data, it is also the data publisher's responsibility to clearly communicate any action(s) taken and to indicate if the full data are available upon request. How you generalize sensitive data (for example, restricting the resolution of the data) depends on the species' category of sensitivity. Where there is low risk of adverse outcomes, unrestricted publication of sensitive species data may remain appropriate. See the published guide [Current Best Practices for Generalizing Sensitive Species Occurrence Data](#) or guidance on when and how to generalize or withhold information sensitive biodiversity data [Chapman 2020]. The guide is also available in [French](#) and [Spanish](#).

#### Reporting data generalizations

When generalizing data you should try not to reduce the value of the data for analysis. A clear summary of the data generalization process should be reported for each relevant Event using `dwc:dataGeneralizations`.

For example, if the spatial resolution of locality data for an Event is reduced to the nearest half degree, then `dwc:dataGeneralizations` = `Coordinates generalized from original GPS coordinates to the nearest half degree grid cell` for each Event to which this treatment was applied. If the location information was generalized for every survey site in a nested hierarchy, then at the parent Event level `dwc:dataGeneralizations` = `Coordinates for each event site generalized from original GPS coordinates to the nearest half degree grid cell`.

#### Reporting information withheld

If specific data are not reported with the dataset, a clarifying statement should be provided with each affected Event using `dwc:informationWithheld`.

For example, if sensitive species data are purposefully excluded from the published data, `dwc:informationWithheld` should include a statement along the lines of `Sensitive species occurrence information not reported`.

### 5.5.8. Least specific target category quantity inclusive

The term `eco:isLeastSpecificTargetCategoryQuantityInclusive` indicates if the total number of organisms detected for a `dwc:Taxon` (including all its subgroups) is shown in one record in `dwc:individualCount` or the paried terms `dwc:organismQuantity` and `dwc:organismQuantityType` in the `occurrence` table. This true/false (Boolean) term helps data users know if the numbers given in these terms include all organisms of that `dwc:Taxon`.

- For `eco:isLeastSpecificTargetCategoryQuantityInclusive` to be `true`, the values shared in `dwc:individualCount` or `dwc:organismQuantity` and `dwc:organismQuantityType` for a single Occurrence record are inclusive of all organisms of that `dwc:Taxon` detected during the Event.
- For `eco:isLeastSpecificTargetCategoryQuantityInclusive` to be `false`, the values shared in `dwc:individualCount` or `dwc:organismQuantity` and `dwc:organismQuantityType` for a single Occurrence record are **not** inclusive of all organisms of the `dwc:Taxon` detected during the survey Event. This means that to find the total number of organisms detected for a given `dwc:Taxon`, you need to add up the `dwc:organismQuantity` values from multiple occurrence records within the Event.

See [Guidelines for eco:isLeastSpecificTargetCategoryQuantityInclusive](#) [TDWG Humboldt Extension Task Group 2024] for more information.

### 5.5.9. Verbatim fields

Two verbatim fields are available to provide additional information about an Event.

- **Field notes:** Field notes can be copied, transcribed verbatim, or linked into `dwc:fieldNotes`.
- **Event remarks:** Additional comments about a particular Event that don't fit in any other term can be shared using `dwc:eventRemarks`.

Both fields can be applied to any Event at any level.

*Other survey protocol information and verbatim protocol terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Recommended	eco:hasMaterialSamples	true or false
	eco:hasVouchers	true or false
	eco:isLeastSpecificTargetCategoryQuantityInclusive	true or false
Share if available	eco:materialSampleTypes	wholeOrganism, blood
	eco:voucherInstitutions	AMNH   KUNHM
	dwc:dataGeneralizations	Coordinates generalized from original GPS coordinates to the nearest half degree grid cell, Coordinates for each event site generalized from original GPS coordinates to the nearest half degree grid cell
	dwc:informationWithheld	Sensitive species occurrence information not reported
	dwc:fieldNotes	Notes available in the Grinnell-Miller Library
	dwc:eventRemarks	

## 5.6. Scope and completeness

**Survey scope** identifies the organisms targeted (or not targeted) during a survey. Structured reporting of explicitly stated survey scopes is necessary for evaluating and reporting completeness and is critical to understanding if the data can be used to assert absences (non-detections) of taxa.

**Completeness** indicates the thoroughness of a survey relative to the stated scope. Structured reporting of explicitly stated survey scopes and completeness is necessary for evaluating and reporting completeness and is critical to understanding if the data can be used to assert absences (non-detections) of taxa. Reported scope and completeness information facilitates the ability of downstream data users to interpret species populations, areas of occupancy, infer species absences, etc.

The 'target' and 'excluded' scope terms (e.g., `eco:targetTaxonomicScope`) presented in this section are



the only Event terms designed to capture *intent*. That is, these terms capture the breadth of the information the biological survey *intended* to capture. All other terms should be used to report the actuality of the survey (e.g., what protocol was in practice implemented, what information was actually collected).

## Implementing scope terms

- Scope terms can be applied at any Event level.
- Recommended best practice is to populate scope terms every Event to which they apply. This information should be reported *only* at the Event levels for which the information is explicitly stated; information should not be inferred up or down an Event hierarchy.
- Scope terms of an Event must be populated whenever the scope was in effect to be able to infer absence of detection within that Event whenever the Occurrences linked to that Event do not explicitly state zero counts or when there are no Occurrence records for a given taxon that fell within the taxonomic scope (see Section 3.2.4 Principle of inference in [Properties of hierarchical events in the Humboldt Extension for Ecological Inventories](#)).
- Do *not* retrospectively infer scope terms.

### 5.6.1. Verbatim scope

The complete scope explicitly identifying the full suite of stated parameters defining the breadth of a sampling Event should be reported using `eco:verbatimTargetScope`. `eco:verbatimTargetScope` is particularly useful for capturing scope conditions not covered by existing [taxonomic](#) or [organismal](#) scope terms.

*General scope terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Recommended	<code>eco:verbatimTargetScope</code>	Adult flying insects

### 5.6.2. Taxonomic scope

Reporting taxonomic scope enables reliable, quantitative, and statistical interpretation of survey and monitoring data. Knowledge of taxonomic scope is essential to interpret local non-detection of taxa as local absences. The **taxonomic scope**, stated either as targeted or intentionally excluded taxa, should be reported using `eco:targetTaxonomicScope` and `eco:excludedTaxonomicScope`.

If every organism in the stated `terms:eco[eco:targetTaxonomicScope]` that was observed during an Event was reported, then `eco:isTaxonomicScopeFullyReported = true`; if not, `eco:isTaxonomicScopeFullyReported = false`.

**Knowledge about taxonomic completeness** allows data users to determine how comprehensively an area was sampled.

- If taxonomic completeness is reported,
  - `eco:taxonCompletenessReported = reportedComplete` or `reportedIncomplete` as appropriate and **\*\*the method used to assess completeness reported in `eco:taxonCompletenessProtocols`.**
- If taxonomic completeness is not reported: `eco:taxonCompletenessReported = notReported`.

If a specific person(s) or organization(s) are reported as making the taxonomic identifications relevant

to the stated survey scope(s), they should be acknowledged in `dwc:identifiedBy`. A list of names can be shared with values separated by a `|`. It is not possible to share a list of unique identifiers such as ORCID's at the Event level.

*Taxonomic scope terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Recommended	<code>eco:targetTaxonomicScope</code>	Arthropods
	<code>eco:excludedTaxonomicScope</code>	Aves
Share if available	<code>eco:isTaxonomicScopeFullyReported</code>	true or false
	<code>eco:taxonCompletenessReported</code>	reportedComplete, reportedIncomplete, or notReported
	<code>eco:taxonCompletenessProtocols</code>	Based on sampling effort
	<code>dwc:identifiedBy</code>	

### 5.6.3. Organismal scopes

As with taxonomic scope, providing information about other organismal scopes when relevant enables reliable, quantitative interpretation of survey and monitoring data and can be essential to interpreting local non-detection as local absences. Three categories of terms are available with which to report an explicitly stated target or excluded **organismal scope**, and state whether or not all target organisms observed were reported. Any additional organismal scopes should be reported using `eco:verbatimTargetScope`.

#### Life stage

**Life stage** refers to a distinct phase in an organism's life cycle targeted for or excluded from a survey (see `dwc:lifeStage`). Life stage may represent a specific developmental, growth, and/or reproductive changes in an organism's life.

- If the survey targeted or excluded specific organismal life stages, the information should be reported in `eco:targetLifeStageScope` and `eco:excludedLifeStageScope`.
- If every organism that falling within the life stage scope that was detected during the survey was reported, then `eco:isLifeStageScopeFullyReported` = true. Otherwise, `eco:isLifeStageScopeFullyReported` = false.

A corresponding `dwc:lifeStage` term is available in the Occurrence extension. This term should be used to report life stage information for organism Occurrences in the **occurrence** table.

#### Growth form

**Growth form** refers to the physical characters or habits of an organism, or group of organisms, in a given environment. It describes their specific shape, structure, and/or pattern of construction.

- If the survey targeted or excluded specific organismal growth forms, the information should be reported in `eco:targetGrowthFormScope` and `eco:excludedGrowthFormScope`.
- If every organism falling within the stated growth form scope that was observed during the survey was reported, then `eco:isGrowthFormScopeFullyReported` = true; if not, `eco:isGrowthFormScopeFullyReported` = false.

## Degree of establishment

**Degree of establishment** refers to 'the degree to which an organism survives, reproduces, and expands its range at the given place and time' (see `dwc:degreeOfEstablishment`).

- If the survey targeted or excluded specific organismal degree of establishment, the information should be reported in `eco:targetDegreeOfEstablishmentScope` and `eco:excludedDegreeOfEstablishmentScope`.
- If every organism that was included within the degree of establishment, and was detected during the survey, was reported, then `eco:isDegreeOfEstablishmentScopeFullyReported` = **true**; otherwise, `eco:isDegreeOfEstablishmentScopeFullyReported` = **false**.

A corresponding `dwc:degreeOfEstablishment` term is available in the Occurrence extension. This term can be used to report the degree to which information about establishment is available for organism Occurrences in the **occurrence** table.

*Organismal scope terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Share if available	<code>eco:targetLifeStageScope</code>	<b>larva</b>
	<code>eco:excludedLifeStageScope</code>	<b>adult   juvenile</b>
	<code>eco:isLifeStageScopeFullyReported</code>	<b>true or false</b>
	<code>eco:targetDegreeOfEstablishmentScope</code>	<b>native</b>
	<code>eco:excludedDegreeOfEstablishmentScope</code>	<b>invasive</b>
	<code>eco:isDegreeOfEstablishmentScopeFullyReported</code>	<b>true or false</b>
	<code>eco:targetGrowthFormScope</code>	<b>tree</b>
	<code>eco:excludedGrowthFormScope</code>	<b>shrub</b>

### 5.6.4. Bycatch

**Bycatch** are organisms detected during a survey that were not explicitly targeted in the scope of a study. Bycatch, or a lack thereof, in a dataset can be reported at the taxonomic and organismal levels.

If **taxonomic** bycatch are reported:

- `eco:hasNonTargetTaxa` = **true** for all relevant Events.
- If all **taxonomic** bycatch (`eco:hasNonTargetTaxa` = **true**) captured/observed during an Event are reported in the dataset:
  - `eco:areNonTargetTaxaFullyReported` = **true**, and
  - a list of taxonomic bycatch should be shared in `eco:nonTargetTaxa` using scientific nomenclature. Entries in a list should be separated by a |.

If **organismal** bycatch are reported:

- `eco:hasNonTargetOrganisms` = **true** at all relevant Event levels.

If the dataset does NOT include taxonomic or organismal bycatch:

- `eco:hasNonTargetTaxa` = **false** for all relevant Events and
- `eco:hasNonTargetOrganisms` = **false** for all relevant Events.

*Bycatch terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Share if available	<code>eco:hasNonTargetTaxa</code>	<b>true or false</b>
	<code>eco:areNonTargetTaxaFullyReported</code>	<b>true or false</b>
	<code>eco:nonTargetTaxa</code>	<b>Parabuteo unicinctus   Geranoaetus melanoleucus; Cetoniinae   Aclopiniae   Cyclocephala modesta</b>
	<code>eco:hasNonTargetOrganisms</code>	<b>true or false</b>

### 5.6.5. Habitat scope

If the survey includes an explicitly stated targeted or excluded **habitat scope** these can be reported in `eco:targetHabitatScope` and `eco:excludedHabitatScope`.

The actual habitat observed at a survey site during an Event should be reported in `dwc:habitat`.

*Habitat scope terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Share if available	<code>eco:targetHabitatScope</code>	<b>deciduous forest</b>
	<code>eco:excludedHabitatScope</code>	<b>urban</b>

## 5.7. Sampling Effort

**Sampling effort** communicates information about the likelihood that a type of organism were be detected: greater effort generally equals a higher probability of detection. Clear reporting of sampling effort is necessary for interpretation of measures of completeness and calculation of abundance (relative or absolute) or biomass, and is critical in assessing the ability to compare information and aggregate data across studies.

### Updated recommended best practice

Capture sampling effort information as structured data using the following Humboldt extension terms:

- `eco:isSamplingEffortReported`
- `eco:samplingEffortValue` and `eco:samplingEffortUnit`
- `eco:samplingEffortProtocol`
- `eco:samplingPerformedBy`

The DwC Event term `dwc:samplingEffort` is currently a recommended field when publishing Event datasets to GBIF; however, this term captures sampling effort in an unstructured way. The Humboldt extension includes 5 terms to more explicitly capture different aspects of sampling effort. The

updated **recommended best practice** is to report sampling effort information as structured data using the Humboldt Extension terms. Through these terms, data providers may explicitly provide the following information:

- **Is sampling effort reported?:** Indicate if sampling effort is reported (**true** or **false**) in `eco:isSamplingEffortReported`.
- **Sampling effort protocol:** `eco:samplingEffortProtocol` should contain a textual description of the sampling effort protocol (e.g., number and arrangement of people or sensors deployed, whether or not sensors were mobile or stationary, how frequently observation, measurements, or samples were taken) and/or provide a link to the protocol used.
- **Sampling effort:** report sampling effort (e.g., the total amount of time of the sampling Event, the total number of people involved) value and units (e.g., trap nights, person hours) using the paired terms `eco:samplingEffortValue` and `eco:samplingEffortUnit`.
- **Sampling performed by:** `eco:samplingPerformedBy` should be used to credit the people involved in the sampling event. The names of one or more people can be reported, with individual names in a list separated with `|`. Best practice is to use a unique identifier (e.g., ORCID) if available.
  - NOTE: Because `eco:samplingPerformedBy` has an **IRI** (internationalized resource identifier) equivalent, only a single ORCID can be provided (the term cannot support a list). If more than one ORCID needs to be shared, a list of ORCID's (using the pipe separator between values) can be supplied using the term `dwc:recordedByID` used BUT it must be applied to each relevant Occurrence and located on the **occurrence** table.

*Sampling effort terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Recommended	<code>eco:isSamplingEffortReported</code>	<b>true or false</b>
	<code>eco:samplingEffortProtocol</code>	<b>40 box traps deployed in the afternoon even spacings along 4 parallel 100m transects placed 50m apart and visited after sunrise the next day</b>
	<code>eco:samplingEffortValue</code>	<b>40, 5</b>
	<code>eco:samplingEffortUnit</code>	<b>trap nights, person hours</b>
	<code>dwc:samplingEffort</code>	<b>40 trap nights, 5 person hours</b>
	<code>eco:samplingPerformedBy</code>	<b>A. Townsend Peterson</b>

## 6. Mapping additional survey event information: DwC Extensions

### 6.1. Extended measurement or fact (eMoF) extension

Additional measurements about a site including values and units of measurement and related protocols can be shared for any Event using the extended measurement or fact extension (eMOF). The extension was developed by the [Ocean Biodiversity Information System \(OBIS\)](#), and detailed instructions about implementing the extension are available in the [OBIS manual](#).

Specific information about the terms included in the emof extension (e.g., term names, definitions, and comments) is available in the [GBIF Repository of Schemas](#).



#### Using the **extended measurement or fact (emof)** extension

- Create a new **emof** table
- Add `dwc:eventID` as a column header. The `dwc:eventID` will link the record to the **event** table.
- Add each emof extension field needed as a column header.
- Populate the relevant extension field(s) for each survey Event (`dwc:eventID`) as necessary.

### 6.2. Relevé extension

The **Relevé extension** is designed to capture vegetation plot survey measurements at a survey site. The extension facilitates explicit reporting of:

- The description of the plant community associated the survey
- Aspect and inclination at the survey site
- Percent total cover of all plants and percent cover of trees, shrubs, herbs, cryptograms, mosses, lichens, algae, litter, water, and rocks
- Heights of tree, shrub, and herbaceous layers
- Whether or not mosses or lichens are identified



#### Using the **Relevé extension**

- Create a new **Relevé** table
- Add `dwc:eventID` as a column header. The `dwc:eventID` will link the record to the **Event** table.
- Add each Relevé extension field needed as a column header.
- Populate the relevant extension field(s) for each survey Event (`dwc:eventID`) as necessary. Information for a unique `dwc:eventID` should require only one row in the table.

For an example implementation of the **Relevé extension**, see the example dataset [Vegetation plots collected in dry grasslands throughout Bulgaria and Romanian Dobrudzha](#).

## 7. Mapping Occurrence information



Occurrence information should be saved to a DwC-A **occurrence** file.

Any DwC Event can be associated with one or more Occurrence records. Occurrence information is mapped using the **DwC occurrence extension** and linked to an Event via the appropriate `dwc:eventID`. Each Occurrence record can be mapped to only a single survey Event; however, Occurrence records can be link to any Event level. In the case of a nested hierarchy, the recommended best practice is to link each Occurrence to the lowest possible Event level to maintain specificity. Occurrence information should be contained in the **occurrence** table of the DwC-A.

Each organism Occurrence must include the following information:

- **Event ID (`dwc:eventID`):** Links the Occurrence to the correct Event.
- **Occurrence ID (`dwc:occurrenceID`):** A unique identifier for each Occurrence.
- **Scientific name (`dwc:scientificName`):** The most precise (lowest rank) taxonomic identification of the reported organism(s).
- **Basis of record (`dwc:basisOfRecord`):** The nature of the Occurrence (e.g. human observation, material specimen).

### 7.1. Reporting multiple individuals as a single Occurrence

If multiple individuals of the same taxonomic classification are observed and no additional information about the organisms (e.g., life stage, sex) beyond taxonomic identification is reported, all individuals should be reported as a single Occurrence (e.g., 1 row in the table), with the following information:

- the `dwc:eventID` of the Event when the Occurrence occurred,
- a unique `dwc:occurrenceID`,
- the taxonomic classification of the organisms reported in `dwc:scientificName`, and
- the quantity and unit of organisms observed reported in the paired terms `dwc:organismQuantity` and `dwc:organismQuantityType`.

For example, if four hooded crows (*Corvus cornix*) were observed, a single occurrence with one `dwc:occurrenceID` should be reported. See the table below.

<code>dwc:eventID</code>	<code>dwc:occurrenceID</code>	<code>dwc:basisOfRecord</code>	<code>dwc:scientificName</code>	<code>dwc:organismQuantity</code>	<code>dwc:organismQuantityType</code>
<uniqueEventID>	<uniqueObsID>	HumanObservation	Corvus cornix	4	individuals

### 7.2. Reporting multiple individuals as multiple occurrences

If multiple individuals of the same taxonomic classification are observed and additional information about the organisms (e.g., life stage, sex) is collected, then a unique Occurrence record (row in the **occurrence** table) should be created for each unique combination of taxonomic identification-organism traits.



For example, if 1 adult male and 3 adult females Indian gharials (*Gavialis gangeticus*) were observed alive, two Occurrence records, each with a unique `dwc:occurrenceID` would be reported. See the table below.

<code>dwc:eventID</code>	<code>dwc:occurrenceID</code>	<code>dwc:basisOfRecord</code>	<code>dwc:scientificName</code>	<code>dwc:organismQuantity</code>	<code>dwc:organismQuantityType</code>	<code>dwc:sex</code>	<code>dwc:lifeStage</code>	<code>dwc:vitality</code>
uniqueEventID-1	uniqueObsID-1	HumanObservation	<i>Gavialis gangeticus</i>	1	individuals	male	adult	alive
uniqueEventID-1	uniqueObsID-2	HumanObservation	<i>Gavialis gangeticus</i>	3	individuals	female	adult	alive

The table below outlines the minimum required and recommended terms for each Occurrence, as well some of the more commonly used terms and their recommended usages. However, Darwin Core includes many more terms. It is advisable to take some time to review the [DwC quick reference guide](#) to identify any additional terms that may be able to capture other data reported in the dataset. Sections that may be of particular interest:

- [occurrence](#)
- [identification](#)
- [taxon](#)

*Species occurrence terms, their recommended usage (status), and example data entries.*

Status	Term	Example entry
Required	<code>dwc:eventID</code>	uniqueEventID-1
	<code>dwc:occurrenceID</code>	uniqueObsID-2
	<code>dwc:scientificName</code>	<i>Gavialis gangeticus</i>
	<code>dwc:basisOfRecord</code>	HumanObservation
Recommended	<code>dwc:taxonRank</code>	species
	<code>dwc:kingdom</code>	Animalia
	<code>dwc:organismQuantity</code>	3
	<code>dwc:organismQuantityType</code>	individuals
	<code>dwc:occurrenceStatus</code>	present or absent
Share if available	<code>dwc:vernacularName</code>	Indian gharial
	<code>dwc:sex</code>	female
	<code>dwc:lifeStage</code>	adult
	<code>dwc:establishmentMeans</code>	native
	<code>dwc:degreeOfEstablishment</code>	native
	<code>dwc:vitality</code>	alive

## 7.3. Reporting absences

**Absences** are defined here as the lack of detection of organisms that are explicitly stated to be part of the target taxonomic scope. Information regarding the absence of detection of a type of taxon can

be reported explicitly or implicitly within a DwC-A.

The reporting of absences only provides meaningful information when taxonomic scope is fully reported (`eco:isTaxonomicScopeFullyReported = true`).

If taxonomic scope is not fully reported (`eco:isTaxonomicScopeFullyReported = false`), Occurrence records of zero individuals are uninterpretable because the user of these data cannot know whether there are any taxa that were not detected but were not reported with Occurrence records of zero individuals. Implicit reporting of absences is impossible if taxonomic scope is not fully reported (`eco:isTaxonomicScopeFullyReported = false`).

**If absences are implicitly reported**, each end user of data will need to reconstruct explicit absences for themselves for each taxon of interest to them because any unreported taxon (i.e. any taxon without an Occurrence record) is known to be absent. Again, this will only be possible if a taxonomic scope is fully reported (`eco:isTaxonomicScopeFullyReported = true`).

### Explicit documentation of absences

Explicit reporting of absence or non-detection means that the dataset reports, at the Occurrence level (in the `occurrence` table), the lack of detection for each relevant `dwc:Taxon`. When the taxonomic scope is highly constrained, for example being restricted to only one or a few taxa, it is feasible to include Occurrence records for each of the non-detected taxa within the data, with each absence being denoted by reporting `dwc:OccurrenceStatus` as `absent`. To explicitly document taxonomic absences in a DwC-A by including zero-count Occurrence records:

- In the `occurrence` table, for each Occurrence to be reported absent:
  - Populate required fields, e.g., `dwc:eventID` (refer to the [list of required Occurrence terms](#))
  - `dwc:occurrenceStatus = absent`.
  - Recommended best practice is to also populate `dwc:individualCount` and `dwc:organismQuantity` as `0` and `dwc:organismQuantityType` as `Individuals`.
- In the `event` table:
  - If one or more absences are reported at any taxonomic level (`dwc:occurrenceStatus = absent`), then `eco:isAbsenceReported = true` for each relevant Event.
  - A list of absent taxa can be provided using `eco:absentTaxa` at all relevant Events. See the section '[Absence](#)' for details on reporting absence information at the Event level.

Absences should only be reported for taxa within the stated taxonomic and/or organismal scope of a survey. Absence cannot be asserted for bycatch.

### Implicit documentation of absences

Implicit reporting of absence or non-detection, on the otherhand, means that the lack of detection of a `dwc:Taxon` is indirectly suggested through the lack of Occurrence record in the `occurrence` table. When taxonomic scopes are broader, and include hundreds or thousands of species (e.g., a taxonomic scope of a dataset that includes all species of birds in the world), then is not feasible to add occurrence records of zero individuals for all of the species not detected. To implicitly document absences in a DwC-A, `eco:isTaxonomicScopeFullyReported` must be `true` for the Event and either `eco:targetTaxonomicScope` or `eco:excludedTaxonomicScope` must be specified. Then:

- In the `occurrence` table, for each taxon to be implicitly reported absent, there will not be any Occurrence record created.
- In the `event` table, `eco:isAbsenceReported = false` for all relevant Events because no absences are

explicitly reported. See the section '[Absence](#)' for details on reporting absence information at the Event level.

Absences should only be reported for taxa within the stated taxonomic and/or organismal scope of a survey. Absence cannot be asserted for bycatch.

*Terms to indicate absence, or non-detection, and their recommendation usage.*

Table	Recommended usage	Term	Example entry
Occurrence	Required	dwc:occurrenceStatus	present or absent
Occurrence	Recommended	dwc:individualCount	3
Occurrence		dwc:organismQuantity	3
Occurrence		dwc:organismQuantityType	individuals
Event		eco:isAbsenceReported	true or false
Event		eco:isTaxonomicScopeFullyReported	true or false
Event	Share if available	eco:absentTaxa	

## 7.4. Reporting abundances

To capture [abundance](#) in a dataset or at a specific Event level:

- In the [occurrence](#) table,
  - Populate required fields, e.g., dwc:eventID (refer to the [list of required Occurrence terms](#))
  - use the paired terms dwc:organismQuantity and dwc:organismQuantityType to report the observed abundance for each reported Occurrence. For example, if 3 individuals of a species were observed, dwc:organismQuantity = 3 and dwc:organismQuantityType = Individuals.
- In the [event](#) table:
  - The inclusion of abundance information in the dataset, even if this information is not reported for all taxa, should be indicated by populating eco:isAbundanceReported as true.
  - The existence of an abundance cap should be captured using the Boolean term eco:isAbundanceCapReported and the value of that cap reported in eco:abundanceCap.

If the dataset or relevant Event does not include abundance information, then it is recommended that the following terms be populated as follows in the [event](#) table at the appropriate level(s) within the Event hierarchy:

- eco:isAbundanceReported = false
- eco:isAbundanceCapReported = false

See the section '[Abundance](#)' for details on reporting abundance information at the Event level.

*Terms to indicate abundance, the table on which they should be provided, and their recommendation usage.*

Table	Recommended usage	Term	Example entry
Occurrence	Recommended	dwc:individualCount	3
Occurrence		dwc:organismQuantity	3
Occurrence		dwc:organismQuantityType	individuals
Event		eco:isAbundanceReported	true or false
Event		eco:isAbundanceCapReported	true or false
Event	Share if available	ecoabundanceCap	5

## 7.5. Capturing species co-occurrence and species interactions

The **resource relationship extension** can be used to link information related across multiple Occurrences (may be from the same or from different Events), such as:

### An Occurrence with another Occurrence

The table below highlights an example from the dataset **Potential host plant records recovered from ECOAB wild bee collection, Mexico** published by *Comisión nacional para el conocimiento y uso de la biodiversidad*. In this example, a *Bombus ephippiatus* bee visits a species of runner bean, *Phaseolus coccineus*.

Table	Recommended usage	Term	Example entry
Occurrence	Required	dwc:occurrenceID	ECOSUR-SC:ECOAB:861
ResourceRelationship		dwc:resourceID	ECOSUR-SC:ECOAB:861
ResourceRelationship		dwc:relatedResourceID	PHACOC
ResourceRelationship	Recommended	dwc:relationshipOfResource	visits flowers of
ResourceRelationship		dwc:relationshipOfResourceID	<a href="http://purl.obolibrary.org/obo/RO_0002622">http://purl.obolibrary.org/obo/RO_0002622</a>

### An Occurrence with a material sample

The table below highlights an example from the dataset **University of Michigan Museum of Zoology, Division of Reptiles & Amphibians** published by University of Michigan Museum of Zoology. In this example, a skin sample from a female toad of *Bufo americanus* is preserved at the University of Michigan Museum of Zoology along with other body parts.

Table	Recommended usage	Term	Example entry
Occurrence	Required	dwc:occurrenceID	3df9bda0-c41a-4130-83bf-8603ae9c22bb
ResourceRelationship		dwc:resourceID	3df9bda0-c41a-4130-83bf-8603ae9c22bb
ResourceRelationship		dwc:relatedResourceID	urn:catalog:UMMZ:Herps:22

Table	Recommended usage	Term	Example entry
ResourceRelationship	Recommended	dwc:relationshipEstablishedDate	2019-01-14
ResourceRelationship		dwc:relationshipAccordingTo	VertNet

## 8. Specific biological survey types

This section is intended to help data publishers identify available resources that enable sharing of some specific types of biological survey data through GBIF.

### 8.1. Camera trap survey data

Refer to [Best Practices for Managing and Publishing Camera Trap Data \[Reyserhove et al. 2023\]](#) for help in standardizing and publishing camera trap data.

An R package, [camtrapDP \[Bubnicki et al. 2024\]](#), is available to read and restructure camera trap data into Darwin Core. NOTE: The camtrapDP package currently only transforms data into occurrence core format but is nonetheless useful in structuring species occurrences derived from camera trap data into a Darwin Core Archive.

### 8.2. DNA and metabarcoding data

The [DNA derived data extension](#) includes terms that will be of use. For more specific guidance in standardizing and publishing DNA sequence and metabarcoding data, refer to [Publishing DNA-derived data through biodiversity data platform \[Abarenkov et al. 2023\]](#). The guide is available in [French](#), [Spanish](#), and [Chinese](#) in addition to English.

The GBIF Metabarcoding Data Toolkit (MDT) is a useful resource. Learn more about GBIF's [Metabarcoding Programme \(MDP\)](#).

### 8.3. Environmental impact assessments

Refer to [Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments \[GBIF Secretariat & IAIA 2020\]](#) for help with sharing primary biodiversity data resulting from environmental impact assessments. The guide is also available in [French](#) and [Spanish](#).

### 8.4. Freshwater biodiversity data

The [Freshwater Data Publishing Guide \[Lento & Schmidt-Kloiber 2025\]](#) supports holders of freshwater biodiversity data by describing best practices and presenting detailed lists of required and recommended data and metadata fields for preparing and sharing such data through GBIF.

### 8.5. Vector-pathogen data

A [guide and data template](#) for disease vector data is available.

### 8.6. Private companies

A [guide](#) is available to help private companies navigate the process of becoming GBIF data publishers [\[Figueira et al. 2020\]](#). The guide is also available in [French](#), [Portuguese](#), and [Spanish](#).

## 9. Additional guidance and reaching out for assistance

Need more information? Check out the following documentation:

- [Humboldt extension for ecological inventories](#)
- [GBIF Technical Documents](#)
- [The OBIS Manual](#)

For any remaining questions, reach out for assistance from:

- The [Humboldt Extension GitHub repository](#): questions about usage, issues with the vocabulary, and recommendations for new terms should be reported as an issue.
- The [GBIF community forum](#).
- The GBIF Node for your country or organization.
  - If your country or organization is a member of GBIF and has an established Node, you can reach out directly to them.
    - If you're uncertain if your country or organization is part of the GBIF network you can [search here](#).
  - If your country or organization is not a member of GBIF, reach out to the [GBIF helpdesk](#) for assistance.
- GBIF help desk
  - In case technical documentation needs improvements, create an issue on the [GitHub tech-docs project](#).
  - Send an email to the [GBIF helpdesk](#).

## 10. Feedback

The authors appreciate every opportunity to improve this guide. If you would like to provide feedback, please do so by submitting a GitHub issue. If you are unfamiliar with this process, refer to the instructions below:

- Create a GitHub account (see [video how-to](#)).
- If you see something, say something, by creating or commenting on issues on GitHub (see [video how-to](#)). Please refer to specific sections or lines in your recommendations.

Please remember that all interactions within this process must adhere to the GBIF Code of Conduct, which aims to encourage a "safe, hospitable, and productive environment" that is "professional, respectful and harassment-free for all participating."



# Glossary

**absence**

the lack of detection of organisms explicitly stated as belonging to a target taxonomic scope.

**abundance**

the number of individuals of the same taxonomic designation in a specific area at a specific time.

**biological or biodiversity survey**

a systematic effort to collect information about the biological organisms of a specific area at a given time.

**bycatch**

organisms detected during a survey that were not explicitly targeted in the survey scope.

**child Event**

a child Event is any dwc:Event that is contained entirely within a single parent Event.

**compilation**

summary inventory resulting from the combination of information from multiple existing sources (as described by Guralnick et al. 2018), which may be compiled from other data sources and literature searches. Compilations are aggregations of multiple studies, and may combine surveys employing different protocols, processes, and observers, often with variable reporting of the methods employed.

**completeness**

an indication of the thoroughness of a survey relative to the stated scope.

**controlled vocabulary**

a list of accepted values that can be used for a term.

**Darwin Core standard - Dwc**

a standard for sharing and publishing biodiversity data, originating from the Biodiversity Information Standards (TDWG) community. In principle, a set of terms used for describing different components of biodiversity observations, such as sampling events, occurrences and taxa. Current Darwin Core terms are described in the Darwin Core Quick Reference Guide.

**Darwin Core Archive - Dwc-A**

compressed (ZIP) file format for exchange of biodiversity data compiled in accordance with the Darwin Core (Dwc) standard. Essentially a self-contained set of interconnected CSV files and an XML document describing included files and data columns, and their mutual relationships.

**data mapping**

the process of matching fields from one database to another.

**degree of establishment**

the degree to which an organism survives, reproduces, and expands its range at the given place and time (see `dwc:degreeOfEstablishment`).

**Digital object identifier - DOI**

long-lasting reference used to uniquely identify (and locate) digital information objects, such as a biodiversity data set or a scientific publication.

**ecological monitoring**

the collection of information about the state of a system in the natural world through repeated surveys.

**event**

an action that occurs at some location during some time (see `dwc:Event`).

**FAIR data**

data that meet the FAIR principles of *\*F*indability, *\*A*ccessibility, *\*I*nteroperability, and *\*R*eusability. Refer to <https://www.go-fair.org/fair-principles/>.

**growth form**

the specific shape, structure, and/or pattern of construction of an organism or group of organisms.

**Humboldt extension for ecological inventories**

a vocabulary extension to the Darwin Core Event class aimed at capturing detailed data on sampling context (e.g., survey protocols, scopes, and effort) in a structured manner. See [Humboldt Extension for Ecological Inventories](#).

**Internationalized resource identifier (IRI)**

an internet protocol standard that facilitates the identification of online resources. It builds on the Uniform Resource Identifier (URI) protocol by expanding the set of permitted characters beyond ASCII. See more at <https://www.w3.org/International/O-URL-and-ident.html>.

**life stage**

refers to a distinct phase in an organism's life cycle. This may represent a specific developmental, growth, and/or reproductive changes in an organism's life.

**material sample**

an entity '... that represents an entity of interest in whole or in part' (`dwc:MaterialSample`). Essentially all material samples are physical specimens collected during a survey Event.

**nested dataset**

a complex survey dataset consisting of multiple related Event levels represented explicitly in a hierarchical (i.e. nested) structure by creating higher-level parent Events.

**non-nested dataset**

a simple survey dataset consisting of a single sampling Event level.

**occurrence**

an existence of an Organism (sensu `dwc:Organism`) at a specific place at a specific time.

**open data**

data that can be freely used, re-used, and redistributed by anyone.

**paired terms**

mutually interdependent sets of terms that must be populated together for complete information to be present, for example with `eco:eventDurationValue` and `eco:eventDurationUnit`.

**parent Event**

any `dwc:Event` whose `dwc:eventID` is a `dwc:parentEventID` for at least one other `dwc:Event`.

**sampling effort**

aspects of observer behaviour that can vary from one sampling event to another, and which influence the probability that an organism will be detected given that the organism is present.

**sampling Event data**

structured information that describes the broader context surrounding the detection (or non-detection) of an organism in a specific time and place, including documentation of sampling protocol and sampling effort (see definitions for these terms in this Glossary). Sampling Event data encompasses species occurrences, material samples (such as whole or partial specimens), genetic sequences, multimedia, etc. Sampling Event data are typically quantitative and follow documented protocols resulting from sampling Events such as biological inventories, systematic monitoring surveys, and collecting expeditions, as well as structured citizen science efforts. These data can range in complexity from very simple—a single event with a single occurrence or no occurrences—to hierarchically complex, with multiple layers of parent-child Events and any combination of accompanying data types (occurrences, material samples, etc.).

**sampling Event hierarchy**

the description of a survey's sampling design as a series of Events using Darwin Core terms.

**sampling protocol**

details of how a survey was conducted, capturing the sequence of steps and aim to supply a user with information about how the data were acquired and are applicable elsewhere.

**scope**

a description of the restrictions placed on the range of types of organisms being targeted (or not targeted) during a survey, such as the range of species or ages.

**site**

the location at which observations are made or samples and/or measurements are taken. The configuration of an event site can vary in configuration from a point in space to a line to an area to a volume.

**survey design**

the pre-determined constraints of a sampling strategy, including how the survey Event sites (e.g., stations, plots, transects) are laid out, temporal, methodological, etc..

**voucher**

a physical specimen or other material sample collected and accessioned into a museum collection in support of a specific project or survey effort.

# References

- [Abarenkov et al. 2023] Abarenkov K, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Jeppesen TS, Kõljalg U, Lundin D, Nilsson RN, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C, and TG Frøslev. 2023. Publishing DNA-derived data through biodiversity data platforms, v1.3. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-vf1a-nr22>.
- [Bubnicki et al. 2024] Bubnicki JW, Norton B, Baskauf SJ, Bruce T, Cagnacci F, Casaer J, Churski M, Cromsigt JPGM, Farra SD, Fiderer C, Forrester TD, Hendry H, Heurich M, Hofmeester TR, Jansen PA, Kays R, Kuijper DPJ, Liefing Y, Linnell JDC, Luskin MS, Mann C, Milotic T, Newman P, Niedballa J, Oldoni D, Ossi F, Robertson T, Rovero F, Rowcliffe M, Seidenari L, Stachowicz I, Stowell D, Tobler MW, Wieczorek J, Zimmermann, F, and P Desmet. 2024. Camtrap DP: an open standard for the FAIR exchange and archiving of camera trap data. *Remote Sensing in Ecology and Conservation*, 10:283-295. <https://doi.org/10.1002/rse2.374>.
- [Campbell et al. 2021] Campbell I, Behrens K, Hesse C, and Chaon P. Habitats of the World: A Field Guide for Birders, Naturalists, and Ecologists, Princeton: Princeton University Press, 2021. <https://doi.org/10.1515/9780691225968>.
- [Chapman 2020] Chapman AD. 2020. Current best practices for generalizing sensitive species occurrence data. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc-5jp4-5g10>.
- [De Pooter et al. 2017] De Pooter D, Appeltans W, Bailly N, Bristol S, Deneudt K, Eliezer M, Fujioka E, Giorgetti A, Goldstein P, Lewis M, Lipizer M, Mackay K, Marin M, Moncoiffé G, Nikolopoulou S, Provoost P, Rauch S, Roubicek A, Torres C, van de Putte A, Vandepitte L, Vanhoorne B, Vinci M, Wambiji N, Watts D, Klein Salas E, and F Hernandez. 2017. Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. *Biodiversity Data Journal*, 5:e10989. <https://doi.org/10.3897/BDJ.5.e10989>.
- [Dimaki & Legakis 1999] Dimaki M and A Legakis. 1999. The reptile fauna of the Fourni Archipelago (Eastern Aegean, Greece). *Herpetozoa*, 12(3/4), 129-133.
- [Figueira et al. 2020] Figueira R, Beja P, Villaverde C, Vega M, Cezón K, Messina T, Archambeau A, Johaadien R, Endresen D, and D Escobar. 2020. Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-b8hq-me03>.
- [GBIF 2011] GBIF (2011). GBIF Metadata Profile – How-to Guide, (contributed by ÓTuama, Éamonn, Braak K, and D Remsen), Copenhagen: GBIF Secretariat ISBN:87-92020-24-0, accessible online at: <https://ipt.gbif.org/manual/en/ipt/3.0/gbif-metadata-profile>.
- [GBIF 2018] GBIF (2018) Best practices in publishing sampling-event data, version 2.2. Copenhagen: GBIF Secretariat. <https://ipt.gbif.org/manual/en/ipt/3.0/best-practices-sampling-event-data>.
- [GBIF Secretariat & IAIA 2020] GBIF Secretariat & IAIA. 2020. Best practices for publishing biodiversity data from environmental impact assessments. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-5xdm-8762>.
- [Guralnick et al 2018] Guralnick R, Walls R, and W Jetz. 2018. Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography*, 41:713-725. <https://doi.org/10.1111/ecog.02942>.
- [Heberling et al. 2021] Heberling JM, Miller JT, Noesgaard D, Weingart SB, and D Schigel. 2021. Data integration enables global biodiversity synthesis. *PNAS*, 118(6):e2018093118. <https://doi.org/10.1073/pnas.2018093118>.
- [Ingenloff 2025] Ingenloff K. 2025. Survey and Monitoring Data Quick-Start Guide: A how-to for updating a Darwin Core dataset using the Humboldt Extension. GBIF Secretariat: Copenhagen.

<https://doi.org/10.35035/doc-7t3p-ve38>.

- [IPBES 2019] IPBES. 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science–Policy Platform on Biodiversity and Ecosystem Services. ES Brondizio, J Settele, S Díaz, and HT Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages. <https://doi.org/10.5281/zenodo.3831673>.
- [Keith et al. 2020] Keith DA, Ferrer-Paris, J.R., Nicholson, E, and Kingsford, R.T. (eds.). 2020. The IUCN Global Ecosystem Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Gland, Switzerland: IUCN.
- [Lapatas et al. 2015] Lapatas V, Stefanidakis M, Jimenez RC, Via A, and MV Schneider. 2015. Data integration in biological research: an overview. *Journal of Biological Research*, 22(1):9. <https://doi.org/10.1186/s40709-015-0032-5>.
- [Lento & Schmidt-Kloiber 2025] Lento J & A Schmidt-Kloiber. 2025. Freshwater data publishing guide. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-sw3k-w725>.
- [Leonelli 2016] Leonelli S. 2016. Data-Centric Biology: A Philosophical Study, Chicago: University of Chicago Press, 2016. <https://doi.org/10.7208/9780226416502>.
- [NEON 2025] NEON (National Ecological Observatory Network) [1]. 2025. NEON Tick pathogen status (DP1.10092.001), RELEASE-2025. <https://doi.org/10.48443/8nhe-cp13>. Dataset accessed from <https://data.neonscience.org/data-products/DP1.10092.001/RELEASE-2025> on xxx.
- [NEON 2025] NEON (National Ecological Observatory Network) [2]. 2025. NEON Ticks sampled using drag cloths (DP1.10093.001), RELEASE-2025. <https://doi.org/10.48443/6zpz-5z19>. Dataset accessed from <https://data.neonscience.org/data-products/DP1.10093.001/RELEASE-2025> on xxx.
- [Reyserhove et al. 2023] Reyserhove L, Norton B, and P Desmet. 2023. Best practices for managing and publishing camera trap data. GBIF Secretariat: Copenhagen. <https://doi.org/10.35035/doc-0qzp-2x37>.
- [Sampling event Data] Sampling event data. <https://ipt.gbif.org/manual/en/ipt/latest/sampling-event-data>.
- [TDWG Humboldt Extension Task Group 2024] TDWG Humboldt Extension Task Group [1]. 2024. isLeastSpecificTargetCategoryQuantityInclusive Guidelines. Biodiversity Information Standards (TDWG). <http://rs.tdwg.org/dwc/doc/inclusive/2024-02-28>.
- [TDWG Humboldt Extension Task Group 2024] TDWG Humboldt Extension Task Group [2]. 2024. Humboldt Extension vocabulary list of terms. Biodiversity Information Standards (TDWG). <http://rs.tdwg.org/dwc/doc/eco/2024-03-26>.
- [TDWG Humboldt Extension Task Group 2024] TDWG Humboldt Extension Task Group [3]. 2024. Properties of hierarchical events in the Humboldt Extension for Ecological Inventories. Biodiversity Information Standards (TDWG). <https://eco.tdwg.org/hierarchy/>.
- [Thorpe et al. 2016] Thorpe ASDT, Barnett SC, Elmendorf ELS, Hinckley D, Hoekman KD, Jones KE, LeVan CL, Meier LF, Stanish, and KM Thibault. 2016. Introduction to the sampling designs of the National Ecological Observatory Network Terrestrial Observation System. *Ecosphere*, 7(12):e01627. <https://doi.org/10.1002/ecs2.1627>.
- [Wieczorek et al. 2012] Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, and 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 et al. 2016] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A,

Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, and 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>.