Survey and Monitoring Data Quick-Start Guide

A how-to for updating a Darwin Core dataset using the Humboldt Extension

Kate Ingenloff

Version ba1accd, 2025-02-12 17:44:36 UTC

GBIF Global Biodiversity Information Facility

Table of Contents

Colophon

Suggested citation

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

Authors

Kate Ingenloff

Contributors

Cecilie Svenningsen, Javier Gamboa, Yanina V. Sica, Yi Ming Gan, Zachary R. Kachian, Peter Brenton, Wesley M. Hochachka, John Wieczorek, Robert D. Stevenson, Anahita J.N. Kazem, Dmitry Schigel, Steven J. Baskauf & Paula Zermoglio

Licence

Acknowledgement

Survey and Monitoring Data Quick-Start Guide: A how-to for updating a Darwin Core dataset using the Humboldt extension 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-7t3p-ve38

Document control

v1.0, February 2025

Cover image

Summary

The details about a biological survey (how it was carried out, the spatio-temporal scope, the taxonomic groups targeted, who was involved, etc.) are important to properly understand the

structure of the survey and how the published data can be reused. The Humboldt Extension for Ecological Inventories (HE), a vocabulary extension to the Darwin Core (DwC) Event Class, provides a means by which to explicitly report the context in which species occurrence data and/or material specimens were collected. The extension includes 55 terms to capture critical facets of survey design including protocol, scope, and sampling effort in a structured manner, thus enhancing overall FAIRness (specifically findability and interoperability) of biological survey data.

This document will guide GBIF data publishers who (a) already have data formatted as a Darwin Core (DwC) Event dataset through the process of updating their dataset with the Humboldt extension or (b) are comfortable with the DwC Event core and wish to map a new dataset to DwC Event class and Humboldt extension terms.

1. Getting started

The process of updating your Darwin Core Archive (DwC-A) Event dataset with the Humboldt extension will likely involve moving some information from the existing DwC-A metadata to a new `humboldt` table; referring to existing documentation, publications, or associated weblinks related to the dataset; and, when possible, conference with the original data collectors or individuals involved in the design and oversight of the project or survey resulting in the dataset. Data republication efforts described here are expected to increase the value and usefulness of existing event datasets in GBIF, broaden their application and therefore data citation across science and policy reuse scenarios.

Before you get started, we recommend that you prepare by taking the following steps:

1. Review the information already published in your DwC-A for the Event dataset, focusing specifically on the Event, metadata, and extended measurement or facts tables, noting where key information about survey design, sampling protocol, scope, and effort are available.
2. Check the range of existing data citations, including contributions of your dataset to the cited query. Which other reuse avenues are open for the data in question? This may help you to focus on particular data and information elements in the transition to HE.
3. Identify additional dataset resources that can be referred to including supplementary documentation, publications, websites, and dataset contacts and people involved in the data collection or oversight of the project or survey.
4. Review the reported data structure. Does the existing event hierarchy accurately reflect the data and the level of complexity desired? Make necessary changes.
5. Create a `humboldt` table for the DwC-A.

Now, you're ready to capture survey design data using the Humboldt extension following the recommendations below.

2. Updating your DwC Event dataset with the Humboldt extension

Where does a term belong?

- DwC Event terms (preceded by 'dwc', e.g. `dwc:eventID`) should be saved to the `event` table.
- Humboldt Extension terms (preceded by 'eco', e.g. `eco:protocolNames`) should be saved to a separate `Humboldt` table.

Populating terms

Each term in this document is linked with its respective term IRI alias (ex., `eco:protocolNames`). When populating a term in the `event` or `humboldt` table, be sure to refer to the definition, comments, and examples provided in the linked documentation to ensure that you are following recommended usage guidelines.

Refer to Properties of hierarchical events in the Humboldt Extension for Ecological Inventories for guidance in populating Humboldt extension terms across event levels (e.g., from parent event to child event).

2.1. Survey sampling design and event hierarchy

What is survey design and sampling event hierarchy?

Survey sampling design details the sampling strategy and how the survey event sites (e.g. stations, plots, transects) are laid out. The **sampling event hierarchy** is the translation of the survey sampling design into an event-based perspective using Darwin Core terms.

Sampling event hierarchy terms

Historically, only two (2) terms were available to explicitly structure and relate different levels of sampling event hierarchy 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 two (2) terms—`eco:siteCount` and `eco:siteNestingDescription`—to better support complex or nested datasets.

Review your DwC Event dataset to ensure that the survey design is accurately reflected in the use of the five (5) available sampling event hierarchy terms. Where additional events or event levels must be created, be sure to reference A Beginner's Guide to Persistent Identifiers for guidance in creating new persistent identifiers.

Do NOT change existing identifiers if it can be avoided!

2.1.3. Non-nested datasets

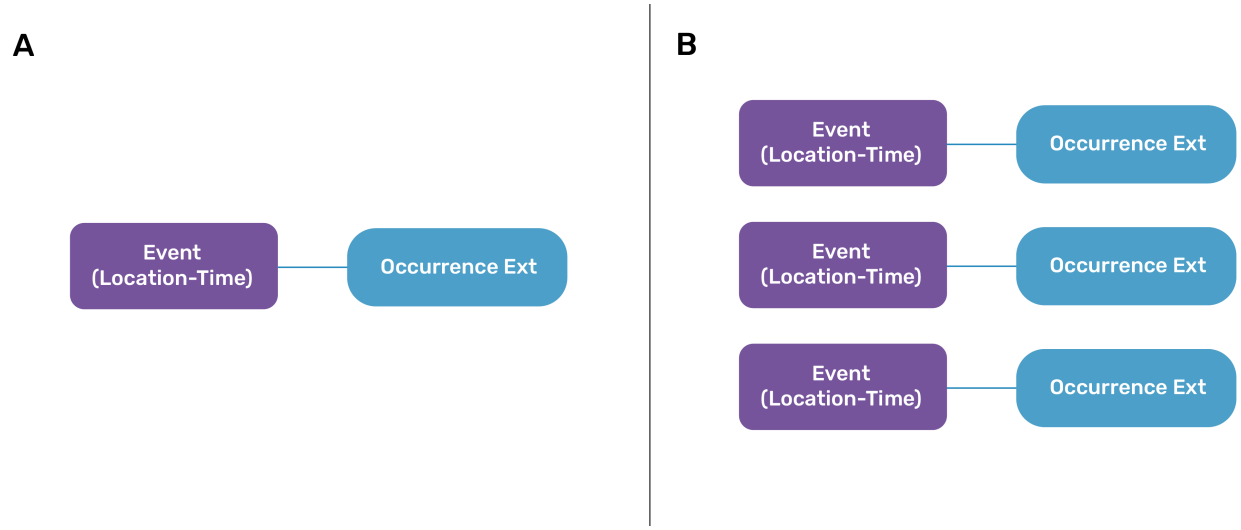

Figure 1. A simple schematic of a non-nested event dataset structure (a) consisting of a single event with associated occurrences related to the event via the occurrence extension and (b) a series of individual events with associated occurrences related to the appropriate event via the occurrence extension.

Non-nested datasets may consist of a single sampling event with a single standardized sampling protocol that is not repeated (Figure 1a) or a series of single sampling events that are not joined by a larger parent event (Figure 1b).

- Each event must have a unique `dwc:eventID`.
- Non-nested datasets will not have a `dwc:parentEventID`.
- `eco:siteNestingDescription` does not need to be populated.

2.1.4. Nested datasets

Nested datasets (multiple nested event levels) are established by relating a child event to a parent event through the child Event's `dwc:parentEventID`. The structure of these datasets can take various forms, but often center either first around the study site second and secondarily on protocol (Figure 2) or conversely, focusing on protocol at higher hierarchical levels and secondarily on locality (Figure 3). Alternatively, time-series dataset are temporally nested datasets (Figure 4).

- Each event must have a unique `dwc:eventID`, and each parent event must have its own `dwc:parentEventID`.
- Nested datasets should, at the parent event level, include the total number of sites sampled in `eco:siteCount` and provide a textual description of the hierarchical sampling design using `eco:siteNestingDescription`.
- If the survey data include a field number for each specific event, this should be shared using `dwc:fieldNumber`.

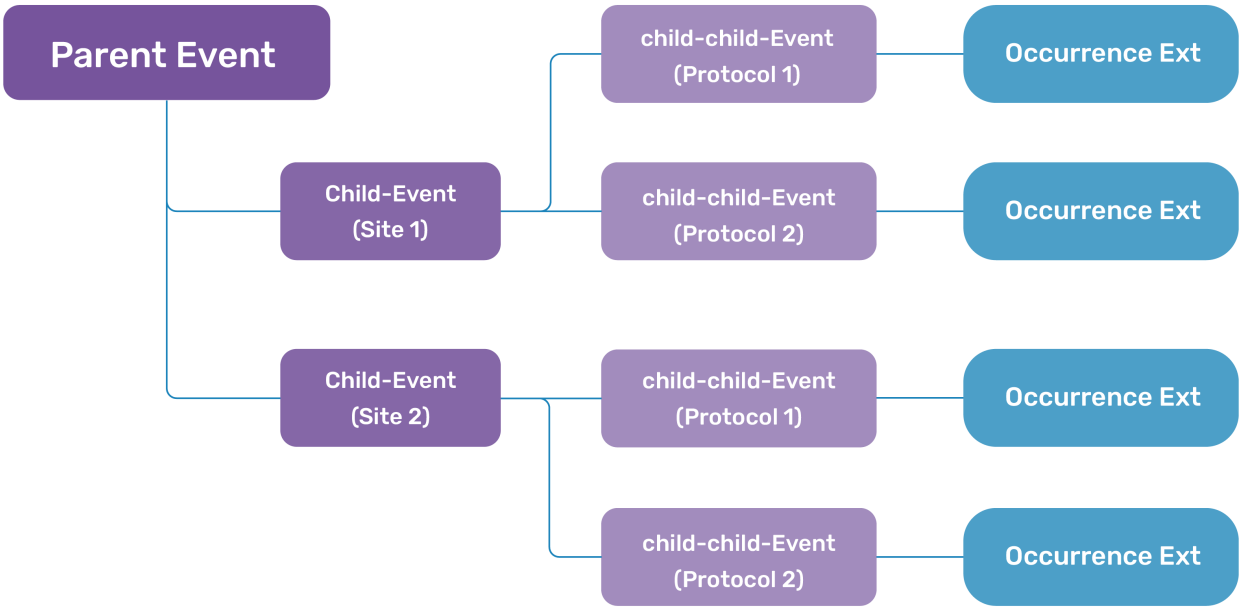

Figure 2. Simplified example schematic of a nested event dataset structure representing a survey (Parent Event, dark red oval) with two survey sites (child-Events, medium red ovals) at each of which two protocols (child-child-Events, light red ovals) are implemented and occurrence information is collected and related to each sampling event using the occurrence extension.

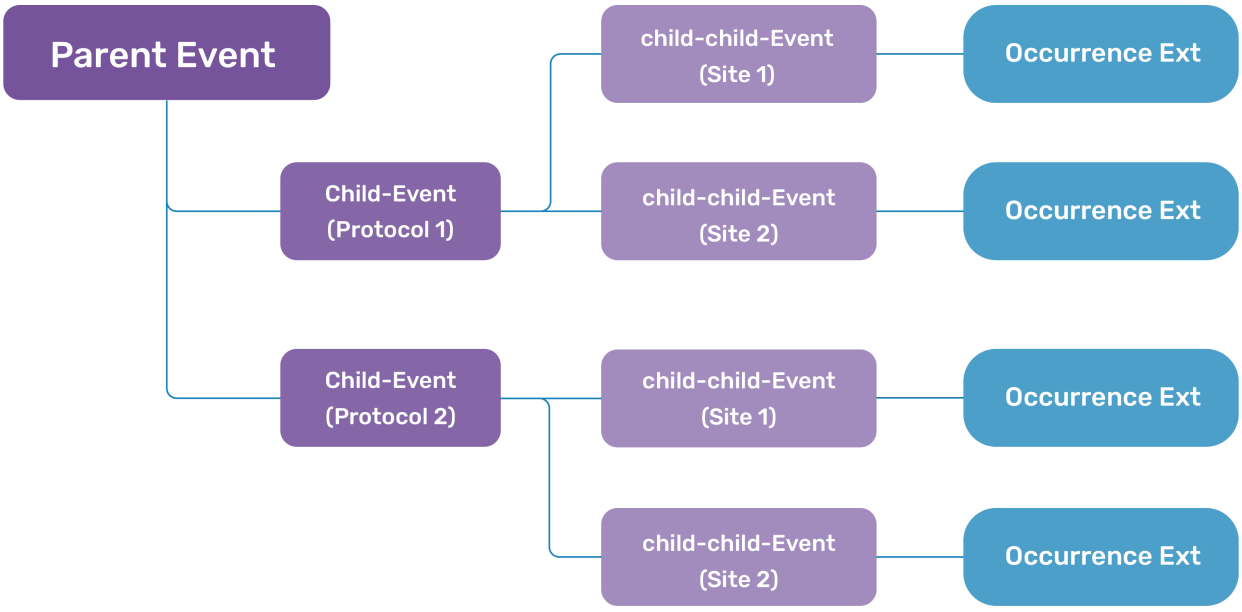

Figure 3. Simplified example schematic of a nested event dataset structure representing a survey (Parent Event, dark red oval) with two protocols (child-Events, medium red ovals) which are each implemented at two survey sites (four independently surveyed child-child-Events, light red ovals) wherein occurrence information is collected in four (likely quantitative) occurrences lists and related to each sampling event using the occurrence extension.

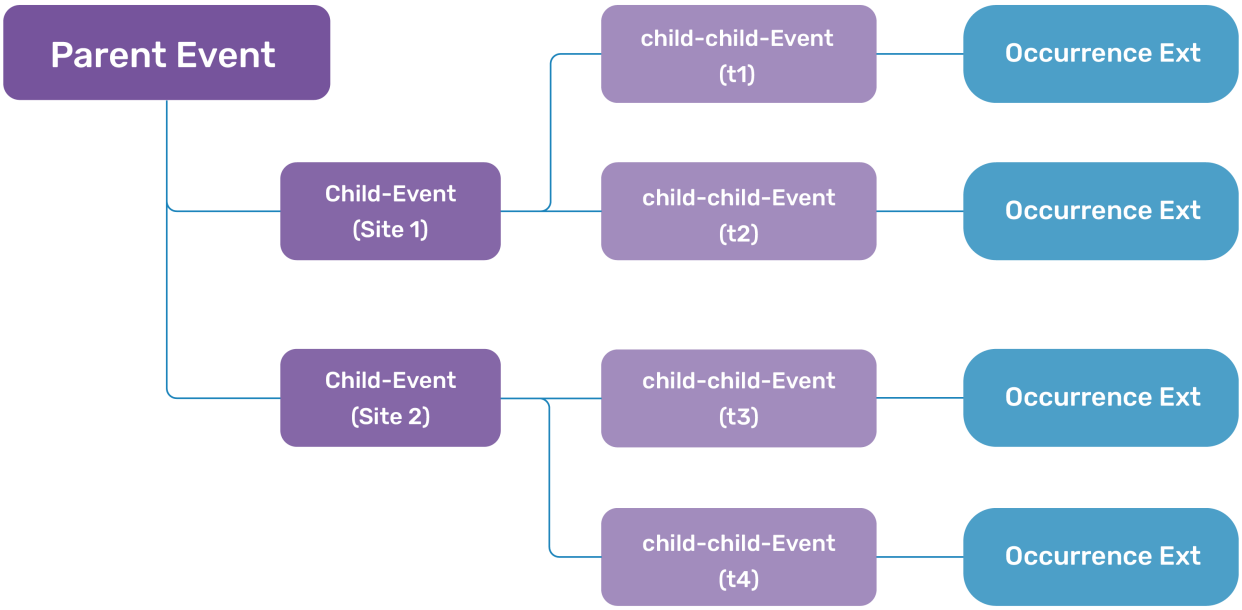

Figure 4. Simplified example schematic of a nested event dataset structure representing a time series survey (Parent Event, dark red oval) with two survey sites (child-Events, medium red ovals) which are each independently sampled at two different times (four independently surveyed child-child-Events, light red ovals) wherein occurrence information is collected in four (likely quantitative) occurrences lists and related to each sampling event using the occurrence extension.

Table 1. Event hierarchy terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Required	`dwc:eventID`	`survey2022_a-2`
Required for nested datasets	`dwc:parentEventID`	`survey2022`
Recommended	`eco:siteCount`	`75`
	`eco:siteNestingDescription`	`25 survey sites each with 3 1m2 quadrats`
Share if available	`dwc:fieldNumber`	`RV Sol 87-03-08`

2.2. Survey event site

Why are site terms important?

An event **site** is a 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 the survey was conducted.

2.2.1. Site description

The following information about a survey event site should be shared for every event level that the information is available:

- **Site names**: report individual sampling event site names using `eco:verbatimSiteNames`. A

concatenated list of site names can be provided at higher event levels.

- **Habitat**: reported habitat at a sampling event site should be recorded in `dwc:habitat`. A concatenated list of habitats can be provided at higher event levels.
- **Weather**: reported weather at a sampling event site should be recorded in `eco:reportedWeather`.
- **Extreme conditions**: reported extreme conditions at a sampling event site at the time of the survey event should be recorded in `eco:reportedExtremeConditions`.
- **Verbatim site description**: verbatim comments (e.g. the original textual description) about a site or sites should be copied in `eco:verbatimSiteDescriptions`.

Table 2. 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: Ammophila arenaria, Betula pendula, Leymus arenarius, Pinus sylvestris`

2.2.2. Site locality

The geographic location and extent of each survey event site should be shared. Historically, five (5) terms were strongly recommended for event datasets in GBIF:

- **Location ID**: a unique identifier for each sampling event site should be shared in `dwc:locationID`.
- **Country code**: the two-letter code for the country in which the survey takes place should be provided in `dwc:countryCode`.
- **Latitude-Longitude**: The decimal latitude and longitude and geodetic datum location of the Event site should be provided in `dwc:decimalLatitude`, `dwc:decimalLongitude`, and `dwc:geodeticDatum`.
    - If geographic coordinates are not in decimal lat-long, populate the following fields instead: `dwc:verbatimLatitude`, `dwc:verbatimLongitude`, and `dwc:verbatimCoordinateSystem`.

These terms are still recommended. However, the Humboldt extension includes additional terms providing greater contextual information about the geospatial scope of a sampling event or series of events that should also be included if the information is available.

Survey site area terms

Humboldt extension includes two sets of paired terms by which to report the area of an event or survey site: 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. **Total area sampled** terms (`eco:totalAreaSampledValue` and `eco:totalAreaSampledUnit`) report the total area sampled during an event.

- For **non-nested event datasets**, `eco:geospatialScopeAreaValue` and

`eco:geospatialScopeAreaUnit`, `eco:totalAreaSampledValue` and `eco:totalAreaSampledUnit` may contain the same values.

- In **nested datasets**, `eco:geospatialScopeAreaValue` and `eco:geospatialScopeAreaUnit` should be used to report the full study extent at the parent event level (the total area surveyed in a series of survey Events) and `eco:totalAreaSampledValue` and `eco:totalAreaSampledUnit` should be used to relate the area of child event sites (e.g., individual survey sites). Geospatial scope value(s) should always be greater than or equal to that of total area sampled.

For example, consider the Biowide project which surveyed 130 40x40m plots across Denmark. Here, the project-level parent event would report the full geographic extent of Denmark: `eco:geospatialScopeAreaValue` = `42934` and `eco:geospatialScopeAreaUnit` = `km2`. The associated 130 child events representing each individual survey site would then report the area of the site as `eco:totalAreaSampledValue` = `40` and `eco:totalAreaSampledUnit` = `m2`.

If the sampled unit is NOT an area (such as a filtered volume of water in a zooplankton haul conducted in marine surveys), the paired terms `dwc:sampleSizeValue` and `dwc:sampleSizeUnit` should be used.

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`. While survey site geometry can be provided at any event level, it may be most informative at the parent-most event level in a nested dataset.
- **Verbatim site location information**: A more general text description of the site location, if available, can be shared using `dwc:locality`.

Table 3. Event site geographic locality and scope terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Recommended	dwc:locationID	Trap_138
	dwc:countryCode	SE
	dwc:decimalLatitude	59.3168
	dwc:decimalLongitude	18.0627
	dwc:geodeticDatum	WGS84
	eco:geospatialScopeAreaValue	580000
	eco:geospatialScopeAreaUnit	km2
	eco:totalAreaSampledValue	40
	eco:totalAreaSampledUnit	m2
	dwc:sampleSizeValue	200
	dwc:sampleSizeUnit	m3
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

2.2.3. Vegetation cover

Vegetation cover at a survey event site can be reported in three ways:

- **Verbatim vegetation cover**: verbatim vegetation cover information can be captured in `eco:verbatimSiteDescriptions`.
- **Percent vegetation cover**: simple percent vegetation cover can be recorded as structured data using the extended Measurements or Facts extension (eMoF).
- **Vegetation plot survey**: vegetation plot survey information can be reported using the Relevé extension.

If vegetation cover is reported using one or more of these methods, then `eco:isVegetationCoverReported` = `TRUE`; otherwise, `eco:isVegetationCoverReported` = `FALSE`.

2.3. Survey date and time

Why are survey date and time terms important?

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.

Event date and time terms

- **Event date**: Each event should have a reported date or date range in `dwc:eventDate` regardless of its hierarchical level. Nested datasets should, at the parent event level, report a date range encompassing all survey dates.
- **Event time and duration**: If reported, note the time and duration of each event using `dwc:eventTime` and the paired terms `eco:eventDurationValue` and `eco:eventDurationUnit`.

Refer to GBIF's technical documentation on Date and time interpretation for more guidance on reporting event dates and times.

Table 4. Event date and temporal scope terms, their recommended usage (status), and example data entries.

Status	Term	Example entry
Required	`dwc:eventDate`	`2018-08-29`, `2007-03-01/2008-05-11`
Recommended	`dwc:eventTime`	`08:00Z`
	`eco:eventDurationValue`	`1`
	`eco:eventDurationUnit`	`hour`

2.4. Sampling event protocol

What is sampling protocol?

A **sampling protocol** provides the details of how the sampling was conducted. Clear communication of the sampling protocol implemented is essential to ensuring the reliability, reproducibility, and reusability of a dataset as detailed knowledge of survey methods facilitates data integration and subsequent analysis.

Sampling protocol terms should be populated at every event level possible as inheritance in either direction should not be assumed or inferred between event levels.

2.4.2. Event type

The nature of each sampling event (e.g., survey, inventory, bioblitz) should be reported using `dwc:eventType`. Event type should provide a high level overview of sampling effort type but should not be 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 use that vocabulary. Otherwise, you can refer to the box summarizing common event types below for guidance.

Common event types

- **Project**: Projects are structured initiatives with explicitly stated objective or suite of objectives and with clear targets, timelines, and deliverables. Projects typically are typically linked to non-biological information identifying participating organizations and people (agents), funding agencies, and other high-level administrative information. Biological sampling may only be one facet of a project's scope. 'Project' as an `dwc:eventType` is appropriate only at the highest (parent) 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/organismal scopes, any number of documented sampling protocols, and varying degrees of complexity in survey design. 'Expedition' as a `eventType` is typically most appropriate at higher (parent) event levels in nested dataset.
- **Survey**: A survey is a 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 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 a `dwc:eventType` at lower (child) event levels in nested dataset.
- **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 (child) event levels in nested dataset.

Inventory event types

- If `eventType` = `inventory`, the type of search implemented (e.g., restricted search, open search, opportunistic search, trap or sample, compilation) must be reported in `eco:inventoryTypes`.
- If `inventoryTypes` = `compilation`, the compilation type should be reported using `eco:compilationTypes` and data sources captured using `eco:compilationSourceTypes`.

Table 5. Event type terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Recommended	dwc:eventType	Inventory, Survey, Bioblitz
Recommended if applicable	eco:inventoryTypes	Open search, compilation
	eco:compilationTypes	compilationOfExistingSourcesAndSamplingEvents
	eco:compilationSourceTypes	museumSpecimens, literature

2.4.3. Sampling protocol

samplingProtocol is required to publish an event dataset to GBIF, however the Humboldt extension includes three (3) terms to capture information about sampling protocol in a more structured manner:

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

Updated recommended best practice

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

Table 6. Survey event protocol terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Required	dwc:samplingProtocol	Visual survey

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 2 m 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 www.svampeatlas.dk, and the nomenclature used is in accordance with this database.
	eco:protocolReferences	See Appendix B of Brunbjerg AK, Bruun HH, Brøndum L et al. (2019) A systematic survey of regional multi-taxon biodiversity: evaluating strategies and coverage. BMC Ecol 19: 43. https://doi.org/10.1186/s12898-019-0260-x, https://www.google.com/url?q=https://www.protocols.io/view/nanopore-minion-kxygx3jwkg8j/v1&sa=D&source=docs&ust=1736780486391914&usg=AOvVaw3oB8oSZiV-MKw0Qf1xFZe0

2.4.4. Material samples

What are material samples?

A material sample is an entity "…that represents an entity of interest in whole or in part." Essentially, material samples are specimens collected during the survey event. They may consist of an entire organism, part of an organism, or a genetic sample.

Reporting material samples

If the dataset includes at least one material sample:

- eco:hasMaterialSamples = TRUE at the appropriate child event level and at any relevant parent event level, and
- the type(s) of materials collected should be listed under eco:materialSampleTypes for each

relevant event level

If the dataset or sampling event does not include material samples:

- `eco:hasMaterialSamples` = `FALSE` at all appropriate sampling event levels.

2.4.5. Vouchers

What are vouchers?

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

Reporting vouchers

If the dataset has vouchers:

- `eco:hasVouchers` = `TRUE` at the appropriate child event level and at any relevant parent event level, and
- a list of institutions housing them should be shared in `eco:voucherInstitutions` for each relevant event level.

If the dataset or sampling event does NOT include vouchers:

- `eco:hasVouchers` = `FALSE` at all appropriate sampling event levels

2.4.6. Least specific target category quantity inclusive

The term `eco:isLeastSpecificTargetCategoryQuantityInclusive` provides a means by which to indicate to data users if an organismal occurrence record for a specific event reporting an explicit quantity of that organism via the paired terms `dwc:organismQuantity` and `dwc:organismQuantityType` represents the total number of that organism observed during the event. That is, it answers the question: *is this the only record of that organism during the event?*

- If the quantity reported using these paired terms includes all the organisms of the same taxon sampled/observed in that single occurrence record, then `eco:isLeastSpecificTargetCategoryQuantityInclusive` = `TRUE.`
- If the quantity reported using these paired terms does not include all organisms of the same taxon sampled/observed in that single occurrence record (e.g. there are two or more occurrence records reported for the same event), then `eco:isLeastSpecificTargetCategoryQuantityInclusive` = `FALSE.`

Refer to Guidelines for eco:isLeastSpecificTargetCategoryQuantityInclusive for more information.

2.4.7. Data generalizations & information withheld

Why withhold or generalize information from published biodiversity data?

Although the general recommendation is to share all biodiversity data available at its highest spatio-temporal resolution, situations exist where it is necessary to do so. Refer to *Current Best Practices for Generalizing Sensitive Species Occurrence Data* for guidance on when and how to generalize or withhold information.

Reporting data generalizations

If specific aspects of data within the dataset are generalized, a clear summary of the data generalization process should be reported at the appropriate event level 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 sampling event 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 published dataset, a clarifying statement should be provided at the appropriate event level(s) using the `dwc:informationWithheld`.

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

2.4.8. Verbatim fields

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

- Field notes can be copied, transcribed verbatim, or linked into `dwc:fieldNotes`.
- 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.

Table 7. 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`
	`dwc:informationWithheld`	`Sensitive species occurrence information not reported`
	`dwc:fieldNotes`	`Notes available in the Grinnell-Miller Library`
	`dwc:eventRemarks`	

2.5. Scope and completeness

What are survey scope and survey completeness?

Scope relates to the biodiversity targeted (or not targeted) during a survey. **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.

Scope terms can be applied at any event level and recommended best practice is to report only the information that is explicitly available.

2.5.1. Verbatim scope

The full verbatim 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.

Table 8. General scope terms, their recommended usage (status), and example data entries.

Status	Term	Example entry
Recommended	`eco:verbatimTargetScope`	`Adult flying insects`

2.5.2. Taxonomic scope

Why is taxonomic scope important?

Providing taxonomic scope enables reliable, quantitative, including statistical interpretation of survey and monitoring data. It is essential to interpret local non-detection as local absences.

Taxonomic scope terms

An explicitly stated targeted or intentionally excluded taxonomic scope should be reported using `eco:targetTaxonomicScope` and `eco:excludedTaxonomicScope`.

- If a specific person or persons is recorded as making the taxonomic identifications relevant to the stated survey scope(s), they should be acknowledged via `dwc:identifiedBy`. Best practice is to use a unique identifier (e.g. ORCiD), if available.
- If every organism included in `eco:targetTaxonomicScope` that was observed during an event was reported, then `eco:isTaxonomicScopeFullyReported` = `TRUE;` if not, `eco:isTaxonomicScopeFullyReported` = `FALSE.`

If taxonomic completeness is known, `eco:taxonCompletenessReported` should be populated as either `reportedComplete` or `reportedIncomplete` and the method used to assess completeness reported in `eco:taxonCompletenessProtocols`. If taxonomic completeness is not reported, `eco:taxonCompletenessReported` = `notReported.`

Table 9. Taxonomic scope terms, their recommended usage (status), and example data entries.

Status	Term	Example entry
Recommended	`eco:targetTaxonomicScope`	`Arthropods`
	`eco:excludedTaxonomicScope`	`Aves`, `Mammalia`

Share if available	dwc:identifiedBy	'Kevin Holston', https://orcid.org/0000-0002-9216-2917
	eco:isTaxonomicScopeFullyReported	TRUE or FALSE
	eco:taxonCompletenessReported	reportedComplete, reportedIncomplete, or notReported
	eco:taxonCompletenessProtocols	Based on sampling effort

2.5.3. Organismal scope

Why are organismal scope terms important?

As with taxonomic scope, providing organismal scope information when relevant enables reliable, quantitative interpretation of survey and monitoring data and can be essential to interpreting local non-detection as local absences.

Organismal scope terms

An explicitly stated target or excluded organismal scope, and clarification as to whether or not all target organisms observed were reported, should be indicated using the following terms:

- **Life stage**: eco:targetLifeStageScope, eco:excludedLifeStageScope, eco:isLifeStageScopeFullyReported
- **Growth form**: eco:targetGrowthFormScope, eco:excludedGrowthFormScope, eco:isGrowthFormScopeFullyReported
- **Degree of establishment**: eco:targetDegreeOfEstablishmentScope, eco:excludedDegreeOfEstablishmentScope, eco:isDegreeOfEstablishmentScopeFullyReported

Other organismal scopes should be reported using eco:verbatimTargetScope.

Table 10. Organismal scope terms, their recommended usage (status), and example data entries.

Status	Term	Example entry
Share if available	eco:targetLifeStageScope	larva
	eco:excludedLifeStageScope	adult, juvenile
	eco:isLifeStageScopeFullyReported	TRUE or FALSE
	eco:targetDegreeOfEstablishmentScope	native
	eco:excludedDegreeOfEstablishmentScope	invasive
	eco:isDegreeOfEstablishmentScopeFullyReported	TRUE or FALSE
	eco:targetGrowthFormScope	tree
	eco:excludedGrowthFormScope	shrub

2.5.4. Bycatch

What is bycatch?

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

Bycatch terms

Bycatch can be reported at the taxonomic and organismal levels.

If ***taxonomic*** bycatch information is included in the dataset:

- Populate `eco:hasNonTargetTaxa` as `TRUE` at all relevant event levels.
- If ALL taxonomic bycatch (`eco:asNonTargetTaxa` = `TRUE`) captured/observed during an Event are reported in the dataset, then
  - `eco:areNonTargetTaxaFullyReported` = `TRUE` and
  - a list of taxonomic bycatch should be provided in `eco:nonTargetTaxa`.

If ***organismal*** bycatch are included in the dataset, then

- `eco:hasNonTargetOrganisms` = `TRUE` at all relevant event levels.

If the dataset does NOT include taxonomic or organismal bycatch, then at all relevant event levels

- `eco:hasNonTargetTaxa` = 'FALSE' and
- `eco:hasNonTargetOrganisms` = `FALSE`.

Table 11. Bycatch terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Share if available	`eco:hasNonTargetTaxa`	`TRUE` or `FALSE`
	`eco:areNonTargetTaxaFullyReported`	`TRUE` or `FALSE`
	`eco:nonTargetTaxa`	`Parabuteo unicinctus`, `Geranoaetus melanoleucus`; `Cetoniinae`, `Aclopinae`, `Cyclocephala modesta`
	`eco:hasNonTargetOrganisms`	`TRUE` or `FALSE`

2.5.5. Habitat scope

Habitat scope terms

An explicitly stated habitat scope should be reported using `eco:targetHabitatScope` and `eco:excludedHabitatScope`.

Table 12. Habitat scope terms, their recommended usage (status), and example data entries

Status	Term	Example entry
Share if available	`eco:targetHabitatScope`	`deciduous forest`
	`eco:excludedHabitatScope`	`urban`, `grassland`

2.6. Sampling Effort

What is sampling effort?

Sampling effort communicates sampling intensity during a sampling event. Clear reporting of sampling effort is necessary to interpret measures of completeness and calculate abundance (relative or absolute) or biomass and is critical in assessing the ability to compare information and aggregate data across studies.

Sampling effort terms

`dwc:samplingEffort` is strongly recommended to publish `dwc:Event` datasets to GBIF, however, the Humboldt extension includes five (5) terms to more explicitly capture sampling effort information:

- **Is sampling effort reported?**: `eco:isSamplingEffortReported` indicates (`TRUE` or `FALSE`) if sampling effort is reported.
- **Sampling effort**: `eco:samplingEffortValue` and `eco:samplingEffortUnit` report sampling effort value and units (e.g. 4 trap nights).
- **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 performed by**: `eco:samplingPerformedBy` should be used to credit the people involved in the sampling eventSampling effort terms, their recommendation usage, and example data entries. Best practice is to use a unique identifier (e.g., OrcID) if available.

Updated recommended best practice

Capture sampling effort information as structured data using the five (5) Humboldt extension terms:

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

Table 13. Sampling effort terms, their recommended usage (status), and example data entries

Status	Term	Example entry

Recommended	eco:isSamplingEffortReported	TRUE or FALSE
	eco:samplingEffortProtocol	40 box traps deployed in the afternoon even spacings along 4 parallel 100m transects placed 50m apart and visited after sunrise the next day
	eco:samplingEffortValue	40
	eco:samplingEffortUnit	trap nights
	eco:samplingEffort	40 trap nights
	eco:samplingPerformedBy	'A. Townsend Peterson, 'https://orcid.org/0000-0003-0243-2379'

Appendix A: Additional guidance and seeking assistance

Additional DwC Event terms

While all Humboldt extension terms are covered in this guide, the Darwin Core Event terms included are not exhaustive. The full suite of available DwC Event terms that can be applied to a DwC-A Event dataset can be found in the GBIF Repository of Schemas Darwin Core Event page.

Need more information?

Check out the following documentation:

- Humboldt Extension for Ecological Inventories
- GBIF Technical Documents

Or, reach out for assistance from:

- Humboldt Extension GitHub repository: questions about usage, issues with the vocabulary, and recommendations for new terms should be reported as an Issue.
- GBIF community forum on Discourse
- GBIF help desk
  - create an issue on the GitHub tech-docs project
  - send an email