Implementation Experience Report for Humboldt extension for ecological inventories

Authors

People writing the report and participating in the testing

Yanina V. Sica, Wesley M. Hochachka, Yi-Ming Gan, Zachary R. Kachian, Steven J. Baskauf, Robert Stevenson, Kate Ingenloff

Abstract

Access to high-quality ecological data is pivotal to assessing and modeling biodiversity and its change through space and time. Inventory data (i.e., recording multiple species at specific places and times) are particularly relevant to monitoring species' distributions and abundance, but their reliability for use in downstream models depends on reporting the methodology implemented and associated sampling effort and completeness. This critical information about the inventory processes is often either not reported or described in an unstructured manner, greatly limiting potential re-use for other analyses. In order to support the reuse of inventories and to assure better standardization of newly collected data, we developed a vocabulary to standardize inventory data reporting that is designed for broad adoption and use. The Humboldt extension for ecological inventories (eco) was developed as a Darwin Core extension, implemented and tested with real world data. Here we describe its implementation process, discuss its advantages and propose its ratification as a TDWG standard.

Introduction and background

Note: because this introduction provides the rationale for this enhancement to Darwin Core and describes how its necessary features were determined, it serves as the Feature Report required by [Section 4.2.1 of the TDWG Vocabulary Maintenance Specification](#).

While mobilization of incidental point records has been in focus for the broader biodiversity data community for many years, mobilization and sharing of inventory and monitoring efforts data is a much more recent phenomenon. The Darwin Core standard (DwC), although most commonly used to capture data at the individual specimen or observation level, does provide an avenue by which to capture inventory-level data. For instance, DwC contains terms such as dwc:samplingProtocol, dwc:sampleSizeValue, dwc:sampleSizeUnit, dwc:samplingEffort, can currently be used to capture some of the information related to inventories. However, it still provides limited ability to express detailed reporting of the inventory scopes (spatial, temporal, taxonomic, and environmental), the sampling protocol , as well as a whole suite of commonly measured aspects of the inventory processes (e.g., reported measures of sampling effort and completeness).

The lack of a standardized vocabulary by which to characterize biodiversity inventory data is a persistent barrier to their mobilization, integration, and broad re-use. In an effort to overcome these limitations, Guralnick, Walls & Jetz (2018) introduced the Humboldt Core as a proof of concept and demonstrated its implementation in Map of Life. Although originally planned as a new TDWG standard, ratification was not pursued at the time, thus limiting adoption by the broader community.

In 2021 the TDWG Humboldt Task Group was established to review how to best integrate the terms proposed in the original publication with existing standards and implementation schemas. Members of Map of Life, the Global Biodiversity Information Facility (GBIF), VertNet, Atlas of Living Australia (ALA), Ocean Biodiversity Information System (OBIS) and other partners in the larger biodiversity community have been meeting regularly to define a standardized way to accommodate the information needed to describe the inventory process. Since inventories (with or without hierarchical structure) can be considered as Events, it was deemed appropriate to build an extension to the Darwin Core Event Core. Hence, the now renamed **Humboldt Extension for ecological inventories (eco)** aims to include the necessary terms to more fully describe the inventory process. First, the task group revised the original Humboldt Core terms from Guralnick, Walls & Jetz (2018), reformulated definitions, comments, and examples, and discarded redundant terms or added new ones as needed. Second, the vocabulary was implemented in a test instance of the GBIF Integrated Publishing Toolkit (IPT) for users to test their inventory datasets. Here we present three case studies that demonstrate the advantages of using the Humboldt Extension for ecological inventories (eco) in terms of inventory data standardization, sharing, and reuse before it undergoes a public review period as established by the TDWG process.

Through ratification of the Humboldt Extension as a TDWG standard we expect to provide the community with a solution to capture and share inventory data improving biodiversity data discoverability, interoperability, and reusability while lowering the reporting burden. This has been clearly called out as urgently needed in order to meet the new set of goals and targets of the Convention on Biological Diversity's Kunming-Montreal global biodiversity framework.

Development of the vocabulary

Explain how we revised all terms from the original publication and how we defined new terms.

After the Task Group's charter was approved, the Task Group members held weekly meetings for over 2 years to discuss the Humboldt Core framework (Guralnick et al. 2018) and the proposed terms to be included in a standard. We started developing the Humboldt Extension for ecological inventories to provide a standardized vocabulary to report key information on biodiversity inventories in order to maximize the usability and interoperability of these data.

Based on Guralnick, Walls & Jetz (2018), we considered biodiversity inventories as surveys set out to document and identify a particular group of organisms (**taxonomic scope**) in a specific location (**spatial scope**, area of land or a volume of water) over a defined period of time (**temporal scope**) using a specified methodological approach (**protocols**, **sampling design**; see Biological Collections Ontology term to define a taxonomic inventory). Inventories and other monitoring efforts are routinely performed (in space or time) and offer high-quality data for characterizing biodiversity patterns and trends. Inventories have the potential to inform which species co-exist or may be absent in a given geographic space over a period of time, but their usability depends largely on how well they describe the inventory process (Figure 1). For instance, knowledge about sampling protocol and its suitability to address a targeted taxonomic and/or spatiotemporal scope, affects the 'completeness' of the inventory (the proportion of expected species successfully detected) which defines whether an inventory can help inform about potential species absences.

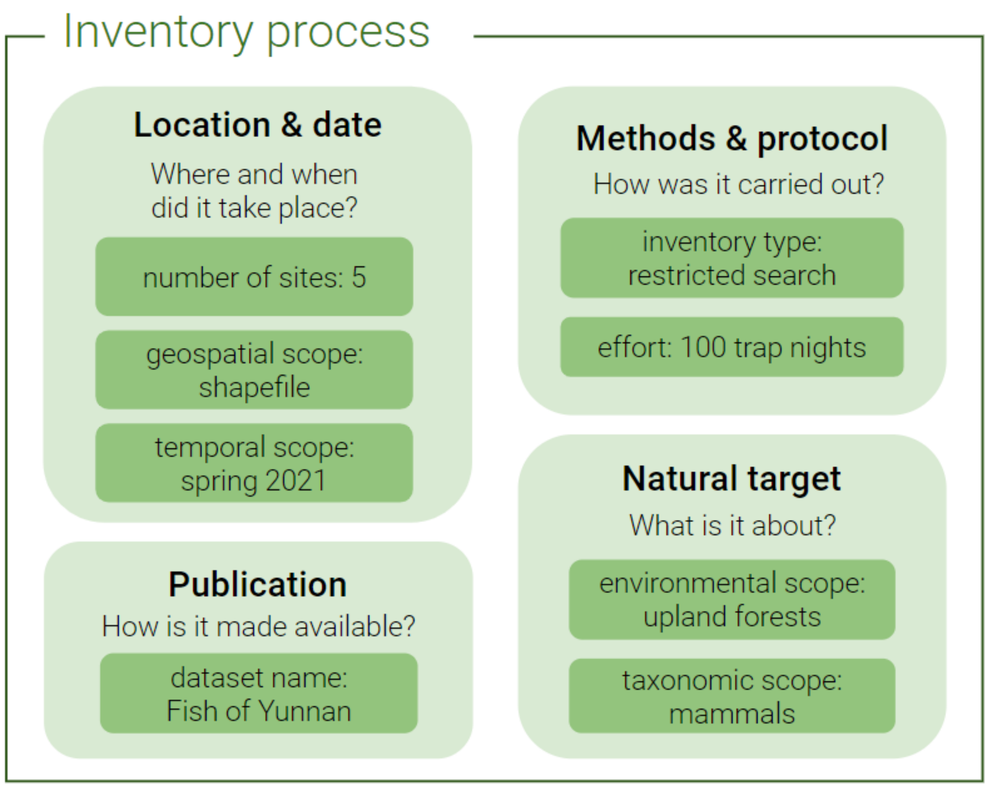

Figure 1. Elements of an inventory process that may produce a species list, with examples in dark green boxes.

In order to capture all elements of the inventory process in a structured manner, we defined six categories (modified from Guralnick et al. 2018; Table 1):

- **General dataset and identification:** terms describing dataset level information.
- **Geospatial scope and habitat**: terms describing where an inventory takes place and the habitat characteristics and environmental conditions of survey sites.
- **Temporal scope**: terms describing when a survey takes place.
- **Taxonomic and organismal scope**: terms describing the target taxonomic group, life stages and/or growth forms, and degree(s) of establishment.
- **Methodology description**: terms describing the inventory methodology including details about inventory type performed, protocol(s) used, abundance and absence reporting, and presence of material samples or vouchers.
- **Completeness and effort**: terms describing inventory completeness and effort.

Within these categories (not formal classes), we revised all original terms, reformulated definitions, and discarded or added new terms where needed. The discussions were documented in a set of meeting notes and the main outcomes were tracked in the Task Group

GitHub repository. We also presented the advances of the Task Group work in TDWG Conferences in 2021 and 2022 (Sica and Zermoglio 2021, Sica et al. 2022) and incorporated new members and feedback from participants. All terms and definitions can be found in the Humboldt Extension Quick Reference Guide.

Table 1. This table outlines the six main categories within the Humboldt Extension used to help identify and describe the inventory process. Terms borrowed from Darwin Core are denoted by "dwc:" prior to the term name (e.g., dwc:identificationReferences). Categories are not defined as formal classes.

Category	Humboldt Extension Terms
General Dataset & Identification	samplingPerformedBy, dwc:identifiedBy, dwc:identificationReferences
Geospatial Scope & Habitat	geospatialScopeAreaInSquareKilometers, totalAreaSampledInSquareKilometers, siteCount, siteNestingDescription, verbatimSiteNames, verbatimSiteDescription, habitatsScope, excludedHabitatScope, reportedWeather, reportedExtremeConditions
Temporal Scope	eventDuration, eventDurationUnit
Taxonomic & organismal Scope	targetTaxonomicScope, excludedTaxonomicScope, targetLifestageScope, excludedLifeStageScope, targetdegreeOfEstablishmentScope, excludedDegreeOfEstablishmentScope, targetGrowthFormScope, excludedGrowthFormScope
Methodology Description	compilationType, compilationSourceTypes, inventoryTypes, protocolNames, protocolDescription, protocolReferences, isAbundanceReported, isAbundanceCapReported, abundanceCap, isVegetationCoverReported, isAbsenceReported, absentTaxa, hasVouchers, voucherInstitutions, hasMaterialSamples, materialSampleTypes, isLeastSpecificTargetCategoryQuantityInclusive

Completeness & Effort	isSamplingEffortReported, samplingEffortProtocol, samplingEffortValue, samplingEffortUnit taxonCompletenessReported, taxonCompletenessProtocol, isTaxonomicScopeComplete, isLifeStageScopeComplete, isDegreeOfEstablishmentScopeComplete, isGrowthFormScopeComplete

Implementation of the vocabulary: testing

Explain how the revised terms were tested using real world data and how the feedback was received.

In a second phase, the larger biodiversity community was able to test the extension with real-world datasets. Several mapping exercises were carried out and a test instance of the GBIF Integrated Publishing Toolkit (IPT) was developed for users to publish their inventory datasets. We also aligned our efforts with other task groups (within and outside TDWG) that are working on developing biodiversity data exchange formats, for instance, the Camera Trap Data Package (CamTrap DP) and the diversified GBIF Data Model). After reporting numerous datasets using the proposed extension, three distinct case studies were selected and presented here to show the broad use of the extension.

Use cases

Use case: eBird Volunteer–Collected Observations of Birds

eBird is a citizen-science project that was started in 2002 for the purpose of collecting, managing and disseminating observations of birds made by bird watchers (Sulllivan et al. 2009, Sullivan et al. 2014). While data were initially collected only in the United States and Canada, the project is now collecting data globally on all species of wild bird species. Most new records are now entered via smartphone apps, although web-based data entry is also still commonly used. The organization coordinating the project and maintaining the database and other software infrastructure is the Cornell Lab of Ornithology, which also has on-staff researchers who use these data for basic and applied research. Because the data are distributed at no cost, they are widely used in academic publications with over 150 publications using these data in 2022 alone (https://science.ebird.org/en/research-and-conservation/publications; accessed 22 March 2023).

There are 6 features to note about eBird's data model:

- All records represent visual or auditory detections of wild birds, and not physical specimens. In some cases there are associated digital "specimens" (still photographs, video, or audio recordings) associated with records. While there is some retrospective correction of identification errors, the vast majority of observation records have their final taxonomic identifications created at the time that records were first entered into the database.
- The data are collected and stored in the form of ***checklists***: groups of records of taxa (mostly species, but also other taxonomic ranks such as subspecies or families). Each checklist can be viewed as a dwc:Event that groups the records of taxa observed by a single group of observers, during a single continuous observation period, and within some contiguous geographical region. This checklist-oriented data model differs from many date compiled during biodiversity inventories, which do not emphasize the storage of records in multi-species sets based on a shared collection/observation event.
- eBird data records for checklists contain a rich array of ancillary information that describes the observation event, and any particular variation in the observation event (e.g., duration of observation event, and distance travelled during the observation event) that will affect the probability of detection/capture of an organism. For example, assume that a species is actually present and that it could have been recorded during an observation event. Then, the likelihood that the species would be recorded will increase with increasing time spent during an observation event, and with increasing distance travelled during the observation event. A very large proportion of variation in numbers of birds reported is the result of variation in observation effort. Thus, these ancillary information describing the observation event is extremely important and needs to be reported in a structured manner for use to control for variation in detection probability when fitting statistical and machine learning models to the data (Johnston et al. 2021).
- The eBird data model contains the concept of a ***complete list***: an indicator of whether a checklist (dwc:Event) contains the records of all species that were detected and identified during the observation event. This is a concept of completeness of reporting; eBird's data model implicitly assumes that the list of species reported is essentially never going to be a complete list of species that were actually present. When a checklist is recorded as being a complete list of the species that were detected and identified, the implication is that unreported species were not detected (*i.e.,* counts of zero individuals can be inferred for all unreported species on complete lists), and therefore the data can be used in analyses that require both presence and absence information. Having this detailed information in a structured manner allows more sophisticated and accurate modeling of responses such as occurrence rates and distributions (Johnston et al. 2021).
- The potential species being reported in any checklist are not constrained, except that reported species should be limited to free-living birds. To speed data entry, a list of all

expected species in an area is provided to observers in the phone apps and on the website. However, any species can be entered on any checklist and bird watchers strive to find and report locally unusual species. Records of unexpected species are flagged for human review, with assessments based on digital media, written descriptions and sometimes also correspondence with the observer(s) reporting the unusual species. The unrestricted taxonomic scope of bird species that are reported means that the inference of zero individuals (see previous point) can apply to any bird species not reported on a list.

- The taxonomy used by eBird contains not only species, but also taxonomic concepts above and below the taxonomic level of species (e.g., genus and sub-species, respectively), as well as visually recognizable hybrids. Within a given species, it is possible to report some individuals at the taxonomic level of a recognizable sub-species and other individuals at the taxonomic level of the full species. The sub-species-level and species-level reports are represented as separate records (i.e. rows of data) within eBird's data model. Because of this, the counts of individuals reported as taxonomic sub-species and taxonomic species must be added together by users of the data in order to ascertain the total number of individual birds reported for that species.

There are many fields in the eBird data model that are not required for general use of the data, or that could be derived from other information stored in the eBird database (e.g., knowing latitude and longitude, it is possible to determine which checklists contain data from any BirdLife International Important Bird area: see http://datazone.birdlife.org/site/mapsearch). Thus, we started by identifying the minimal set of database fields that we believed would contain all of the necessary and sufficient information to estimate distributions, relative abundances and population trends of bird species, based on the experience at the Cornell Lab of Ornithology (Fink et al., 2022). Then we made two tests:

(1) we examined whether each of these "essential" eBird data fields could be represented using a term or terms from the Darwin Core and the proposed Humboldt Extension for ecological inventories, and
(2) we experimented with fitting eBird's "essential" data fields into the current (at the time of the mapping exercise) prototype of GBIF's developing unified model.

From these mapping exercises, we found that most terms perfectly mapped between the two data models. However, we identified a small number of eBird data fields that could not be represented exactly by terms in the tested version of the proposed Humboldt Extension for ecological inventories (plus Darwin Core). These lacks of direct mapping were associated with the concepts of:

- recording information about absence of detection and completeness of reporting,

- the need to sum reported counts from sub-species to obtain the full count of reported individuals for species, and
- describing observation effort.

These concepts were discussed within the TDWG Task group, who deemed appropriate to update some terms of the Humboldt Extension for ecological inventories. Changes done after testing are described in detail in the following sections.

Use case: Field Museum Rapid Inventory Data

Field Museum rapid inventories are cooperative, rigorous surveys of the biological and cultural assets of a priority landscape for conservation. Scientists and long-term residents survey plants, fishes, amphibians, reptiles, birds, and mammals to identify the species, natural resources, and landscape features with high conservation value (at global, national, or local scales), assess the conservation status of those assets, and document threats. The Rapid Inventories team has conducted 31 inventories since 1999, the majority of which (24) have taken place in the Andes-Amazon region. Six were conducted in Cuba and one in China. (See this link for more info on the Rapid Inventories Program).

The data collected are derived from specimens, tissue samples, direct observations, machine observations (camera traps or auditory recordings), or interviews with local experts. Rapid Inventories are designed so that each taxonomic group uses the same locations, except for fishes which use their own locations.

To test the proposed Humboldt Extension for ecological inventories, we took a representative subset of data from a single inventory and applied them to the proposed format. This process involved transcribing data from our internal database to a spreadsheet with the Humboldt Extension terms, where we manually filled in every term that we had data for with the appropriate values. If we were unsure of where to fit the data, we noted that and discussed it with the TDWG Task Group. Most of the data collected in the field mapped perfectly with the proposed extension, the challenges we encountered are described in the following section.

Use case: Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign - data

A marine species inventory dataset in sampling event core, occurrence extension and extended measurement or fact extension was identified by the Antarctic GBIF/OBIS node to be a use case to test with the proposed Humboldt extension for ecological inventories. The dataset consists of catch data collected by Rectangle Mid-water Trawl (RMT) during the BROKE-west cruise in 2006 in Antarctica (Van de Putte et al. 2010).

The dataset was mapped to the testing version of the proposed Humboldt extension terms and published on an Integrated Publishing Toolkit (IPT) test instance: https://ipt.gbif.org/resource?r=brokewest-fish&v=1.5. The advantages and limitations of the Humboldt extension are discussed in detail in the following sections.

Lessons learned

Multiple challenges were identified during the testing phase. In this section we focus on the challenges that we were able to resolve and incorporate in the current extension.

Term descriptions, comments and examples

While testing the proposed Extension with the Field Museum Rapid Inventory data, we came across a few eco:terms that were unclear and needed some improvements in their definitions and comments. Rapid Inventory data also had some data that didn't seem to fit in any of the terms proposed in the Extension. Below, we describe some examples.

The terms "samplingEffortProtocol" and "protocolReference" we weren't sure if this should be a citation for the protocol section of our report that we write for each inventory, or a citation of an external publication. The conclusion in this case was that we could cite both, separated by a "|".

The boolean term "isVegetationCoverReported" also presented some challenges, since we not only have data on weather this was reported but also on the values/metrics for the Events' vegetation cover. It was decided that the values for vegetation cover would fit under "verbatimSiteDescription".

Rapid Inventory data also includes information on moonphase or migration status, as this can affect the sampling. No specific term was added to capture these information. Instead, it was recommended to include this type of information under the Extended Measurement or Fact extension (EMoF). Similarly, we were wondering if it made sense to include a field that addresses conservation status or level of human impact of the area, since some researchers may want to focus on that. It was suggested that this could fit under "habitatScope" or "verbatimLocationDescription".

Lastly, our Rapid Inventories sometimes include data from local or traditional knowledge and we wanted to know where we could include this. It was suggested that "compilationSourceTypes" could be a place for these data.

Data model and Event parent/child structure

When testing the proposed extension with the Field Museum Rapid Inventory data, we had to decide what data model to follow. Specifically, how should we structure the "eventID" and “parentEventID”, since our data could either be organized by location or by protocol. For our inventories, the same locations are used by each taxonomic group (plants, birds, herps, mammals) which each have their own specific sampling protocol. We weren’t sure if we should have each site be a “parent event” with a “child event” for each protocol, or the other way around. After trying out a few different ways of structuring the data, we decided that the best way to do this for our data was by location. This means that each location will have “sub-events” for each protocol that occurred there. Structuring the data this way allows all lower events to inherit the same protocol. This was a very useful yet challenging exercise that needs to be done by any user that want to report and share their biodiversity data. It was decided to improve the documentation on the Humboldt extension for ecological inventories to capture different data structure.

Absence of detection and completeness of reporting

Following eBird case study, it was deemed appropriate to improve the definitions regarding taxon and organism target and completeness and include a term that allows users to indicate whether the dataset contains the records of all species that were detected and identified during the Event. This concept of *completeness of reporting* implies that unreported species were not detected (*i.e.,* counts of zero individuals can be inferred for all unreported species on complete lists), and therefore the data can be used in analyses that require both presence and absence information. Having this detailed information in a structured manner allows more sophisticated and accurate modeling of responses such as occurrence rates and distributions (Johnston et al. 2021). To be able to infer absences, knowing the targeted list of species is not enough, information regarding taxon completeness and reporting completeness is needed.
The task group acknowledges that inferring absences is a persistent issue in ecological problems that requires sophisticated modeling efforts after data collection. Hence, the Humboldt extension working group is collaborating with the Global Biodiversity Information Facility (GBIF) which is exploring a new data model that can accommodate absences of species (access detail information on this project here 14_humboldt_extension_absence_v2).

Inference about organism quantities and abundance

Since the use cases and datasets tested captured different ways of entering information about abundances, it was deemed appropriate to include a term that describes how to get total counts of individuals or organisms in an inventory based on dwc:organismCount.
Add final decision about this term

Use of IRI to avoid long list of objects

Certain terms proposed in the current extension require the recording of multiple objects listed by |. (e.g. eco:samplingPerformedBy, eco:targetTaxonomicScope). In certain cases, the list of expected species to find at a location has been previously published or the list of individuals carrying out a survey is already detailed in some publication. In these cases, by using the IRI version of the terms, links can be used to record such information.

Unresolved issues/ Remaining challenges

Multiple challenges were identified during the testing phase. In this section we focused on the challenges that remain unresolved as they were out of the scope of this Task Group and may need to be discussed in a larger forum.

Paired information of multiple fields of an Event must fit into one record

Due to the limitation of star schema, the Humboldt extension has to be in a 1 to 1 relationship with the core record. In the Antarctic GBIF/OBIS use case example, the Humboldt extension contains multiple fields that contain multiple paired values that extend a single Event. Example of paired values of eco:targetTaxonomicScope and eco:targetLifeStageScope is shown in tables below:

Event Core

eventID	eventRemarks
BROKE_WEST_RMT_006	All life stages of Myctophidae were targeted | Only larvae and juvenile of Macrouridae, Artedidraconidae, Channichthydae and Nototheniidae were targeted by the sampling protocol

Humboldt Extension

eventID	targetTaxonomicScope	targetLifeStageScope
BROKE_WEST		
BROKE_WEST_RMT_006	Myctophidae | Macrouridae | Artedidraconidae | Channichthydae | Nototheniidae	all | larvae and juvenile | larvae and juvenile | larvae and juvenile | larvae and juvenile

This limitation leads to difficulty in interpreting the data as the only way to know the paired information is based on the order in both fields, eco:targetTaxonomicScope and eco:targetLifeStageScope in this example.

Regarding the accommodation of all of eBird's terms describing observation effort, the challenge is that every type of inventory uses a different methodology and the relevant metrics of collection/observation effort will differ across these different methodologies. It would be infeasible for any data model to contain separate terms for every possible descriptor of effort. Some metrics, such as time in collection/observation, are ubiquitous and are represented in Darwin Core + Humboldt Extension, and in the GBIF unified model, as separate terms. Other metrics of effort, such as eBird's number-of-observers metric of effort, will need to be placed into more generic terms that describe both the units of effort and the quantity of effort. Both Darwin Core + Humboldt Extension, and the GBIF unified model, contain appropriate terms to make these generic mappings.

As these challenges are due to the limitation of the current model of Darwin Core Archive, the issue is feedback to Global Biodiversity Information Facility (GBIF) which is developing a new data model that can accommodate such use cases.

The directions in which the information propagates

Some of the terms follow the concept of inheritance. For example, eco:targetTaxonomicScope. It is recommended that in the case of a higher level Event (parent), include all taxonomic groups surveyed in the child Events that contributed to the parent Event. Hence the direction in which the information propagates goes from parent Event to child Event.

Many boolean terms have the directions in which the information propagates in the opposite direction, from child Event to parent Event. For instance, eco:hasMaterialSamples requires the value to be 'true' if the Event is a compilation that contains at least one MaterialSample. If a child Event has the value 'true' for eco:hasMaterialSamples term, then its parent Event must have the value of 'true' for this field.

The task group acknowledges this issue and the solution is to document this as Comments for all of the terms whenever applicable.

Bycatch or non-target species recorded

Any occurrences that are not in eco:taxonTaxonomicScope is assumed to be bycatch. It is implicit instead of explicit.

Other challenges identified in broader testing

Develop a more robust description to identifiedBy ---> organize a meeting with key users and figure out what that definition should look like, then submit to DwC maintenance group (https://github.com/tdwg/dwc/issues/318)

Details of the machine making the observations---> out of scope for this extension. Will need to reach out to Peggy Newman
E.g. IdentificationConfidence (either machine related or not, for machine it is easier because a quantitative value can be generated) → THIS IS NOT AT THE EVENT LEVEL. SHOULD BE SUGGESTED AS A RECOMMENDATION TO ADD IN DWC.
Lists as values → need some use cases for the Technical Architecture Group (TAG) meeting next Monday!

Conclusions

Although some challenges remain, we resolved all the ones that were in our scope. We believe the extension successfully fit all the metadata and dataset description of the test data and propose the extension to go to public review.

References

Fink, D., Auer, T., Strimas-Mackey, M., Lignocki, S., Robinson, O.J., Hochachka, W.M., Jaromczyk, L.O., Rodewald, A.D., Wood, C., Davies, I. & Spencer, A. (2022) eBird Status and Trends, Data Version: 2021; Released: 2022 Cornell Lab of Ornithology. Ithaca, New York. https://doi.org/10.2173/ebirdst.2021

Guralnick, R., Walls, R. and Jetz, W. (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

Johnston, A., Hochachka, W.M., Stimas-Mackey, M., Ruiz-Gutierrez, V., Robinson, O.J., Miller, E.T., Auer, T., Kelling, S.T. & Fink, D. (2021) Analytical guidelines to increase the value of eBird data to estimate species occurrence. Diversity and Distributions, 27, 1265-1277. https://doi.org/10.1111/ddi.13271

Sica YV, Zermoglio PF (2021) Unlocking Inventory Data Capture, Sharing and Reuse: The Humboldt Extension to Darwin Core. Biodiversity Information Science and Standards 5: e74275. https://doi.org/10.3897/biss.5.74275

Sica YV, Ingenloff K, GAN Y-M, Kachian Z, Baskauf SJ, Wieczorek J, Zermoglio PF, Stevenson RD (2022) Application of Humboldt Extension to Real-world Cases. Biodiversity Information Science and Standards 6: e91502. https://doi.org/10.3897/biss.6.91502

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009) eBird: A citizen-based bird observation network in the biological sciences. Biological Conservation, 142, 2282-2292. https://doi.org/10.1016/j.biocon.2009.05.006

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J.W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliff, M.J., Lagoze, C., La Sorte, F.A., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K.V., Trautmann, N.M., Wiggins, A., Van de Putte, A.P., Jackson, G.D., Pakhomov, E., Flores, H., Volckaert, F.A.M. (2010) Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign. Deep Sea Research Part II: Topical Studies in Oceanography, 57(9–10), 956-967. https://doi.org/10.1016/j.dsr2.2008.02.015

Winkler, D.W., Wong, W.-K., Wood, C.L., Yu, J. & Kelling, S. (2014) The eBird enterprise: an integrated approach to development and application of citizen science. Biological Conservation, 169, 31-40. https://doi.org/10.1016/j.biocon.2013.11.003