|  |  |
| --- | --- |
|  | **Biodiversity Information Standards**  **(TDWG)** [http://www.tdwg.org](http://www.tdwg.org/) |

**Humboldt Core Charter, a Task Group of the Observations & specimens Interest Group**

## **Convenor**

* Yanina Sica - Map of Life, Yale University - [yanina.sica@](mailto:yanina.sica@gmail.com)[yale.edu](mailto:walter.jetz@yale.edu)

## **Core Members**

* Walter Jetz - Map of Life, Yale University - [walter.jetz@yale.edu](mailto:walter.jetz@yale.edu)
* Rob Guralnick - Florida Museum of Natural History - [rguralnick@flmnh.ufl.edu](mailto:rguralnick@flmnh.ufl.edu)
* John Wieczorek - VertNet - Darwin Core Maintenance Interest Group - [gtuco.btuco@gmail.com](mailto:gtuco.btuco@gmail.com)
* Paula Zermoglio - VertNet - Darwin Core Maintenance Interest Group - [pzermoglio@gmail.com](mailto:pzermoglio@gmail.com)
* Kate Ingenloff - Map of Life - [kathryn.ingenloff@gmail.com](mailto:kathryn.ingenloff@gmail.com)
* Steve Baskaufs - Vanderbilt University - Darwin Core Maintenance Interest Group - [steve.baskauf@vanderbilt.edu](mailto:steve.baskauf@vanderbilt.edu)
* Tim Robertson - GBIF - [trobertson@gbif.org](mailto:trobertson@gbif.org)
* Dmitry Schiegel - GBIF - dschigel@gbif.org
* Erlend B. Nilsen - Living Norway Ecological Data Network & NINA -e[rlend.nilsen@nina.no](mailto:Erlend.Nilsen@nina.no)
* Peter Brenton - CSIRO, Atlas of Living Australia and TDWG Citizen Science Interest Group and CSA Data & Metadata Working Group - [peter.brenton@csiro.au](mailto:Peter.Brenton@csiro.au)
* Robert D Stevenson - UMass Boston and TDWG Citizen Science Interest Group - [rdstevenson10@gmail.com](mailto:rdstevenson10@gmail.com)

## **Motivation**

Species inventories are routinely performed and offer particular value for characterizing biodiversity and its change. However, reporting standards allowing these inventories to be re‐used, compared to one another, and further integrated with other sources of biodiversity data are lacking, impeding their broadest utility.

The Darwin Core standard currently allows sharing some of the information related to inventories, i.e., terms samplingProtocol, sampleSizeValue, sampleSizeUnit, samplingEffort. However, it still presents a limited ability to express detailed reporting of scope (spatial, temporal, taxonomic, and environmental), as well as a whole suite of commonly measured aspects of inventory sampling processes (e.g., direct or inferred measures of sampling effort).

These limitations could be overcome by formalizing the Humboldt Core, a framework developed to deal specifically with this kind of data (Guralnick et al. 2018), and which is currently implemented in Map of Life (<https://mol.org/humboldtcore/>). While originally planned as a TDWG standard, ratification was not pursued until now. Broader implementation throughout the community has been hampered by a needed review of how to best integrate its terms with existing standards and implementation schemas.

The natural space for the development of such a standard within TDWG is the Observations & Specimens Interest Group. While the purpose of this group is “to explore concepts and methods of biodiversity data description, integration and transfer that fully integrate specimen and observational data into existing data exchange schemas”, they have not yet tackled inventory data, which constitute an interesting mix of Events, Occurrences, and Taxon "checklists".

The purpose of this task group is to enable inventory data to be shared, re-used, compared to one another, and further integrated with other sources of biodiversity data to significantly expand biodiversity dataset discovery, interoperability, and modeling. Such activities have been clearly called out as needed in order to meet the goals of the Convention on Biological Diversity (Schmeller et al. 2015) and promise to lower reporting burden while ensuring clear communication and enhanced value for re-use.

## **Goals Outputs and Outcomes**

* Our **main goal** is to enable biodiversity inventory data sharing as part of the TDWG corpus of standards.
* The **primary deliverable** will be in the form of either:

(a) one or more vocabulary enhancements (extensions) to Darwin Core, or

(b) a new standard for biodiversity inventory data sharing.

Please refer to the Strategy section for clarifications.

* **Other deliverables.** The primary deliverable will be accompanied by:
  + Formal definitions of the terms following the Darwin Core format, adapted from the original definitions provided in Guralnick et al. (2018). No dependencies.
  + A Quick Reference Guide to the terms of the Humboldt Core standard/extension following the Darwin Core Quick Reference Guide model. Depends on term definitions being created.
  + In the case of Darwin Core extension(s), XML document(s) for the extension(s) following the GBIF Darwin Core Extension schema (<http://rs.gbif.org/schema/extension.xsd>). Depends on term definitions being created.
  + In the case of Darwin Core extension(s), publication of at least seven distinct real-world case study data sets via Darwin Core Archives to cover the range of inventory types identified in Guralnick et al. (2018). Depends on the extension XML documents being created.
  + In the case of a new standard, documentation on how to encode the biodiversity inventories in documents of various relevant formats following the model of the Darwin Core Text, XML, and RDF Guides.

It is impossible to predict the duration of some parts of the standards development process, such as the public commentary, in the case of active recommendations for changes. However, several core members of the Task Group have some of their time dedicated to this work. With these considerations in mind, the expectation is to have the outputs ready either for public commentary or for a review manager (depending on standards path determined to be appropriate) by April 2021.

## **Strategy**

* The first task will be to determine whether the full corpus of biodiversity inventory concepts suggested in Guralnick et al. (2018) could be shared in Darwin Core Archives for all of the types of inventories by making extensions that enable any concepts that are currently missing. The output of this exercise would be either a determination that a new standard is needed or a concrete draft of a Darwin Core model to use (which rowtype or Core class, e.g., Event) and the enumeration of extensions that would be needed to enrich the core type.

The hope is that a viable Darwin Core Archive model described above can be defined. If it can not, alternatives will be explored and the work of this task group will be refined and amendments submitted for review.

* Regardless of the outcome of the first task, the definitions provided in the original publication of the Humboldt Core (Guralnick et al. 2018, <https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.02942>) will be formalized using the Darwin Core model. For this work we will also review the formal ontological definitions of inventory concepts that have been developed in the Biological Collections Ontology (BCO, <http://www.obofoundry.org/ontology/bco.html>).
* The development of extensions will follow the process and expected deliverable of the TDWG Vocabulary Maintenance Specification (VMS, <https://github.com/tdwg/vocab/blob/master/vms/maintenance-specification.md>). If a new standard is indicated, development will follow the TDWG Standards Documentation Standard (SDS, <https://github.com/tdwg/vocab/blob/master/sds/documentation-specification.md>) and will be reviewed as required by the TDWG by-laws <https://www.tdwg.org/about/process/> for the ratification of standards.
* Throughout the process we will work closely with the Darwin Core Maintenance Interest Group, some members of whom are also members of this Task Group, and who have experience with the development and incorporation of extensions.
* We will also engage experts to aid with the revision of term definitions as necessary.
* We will work closely with GBIF staff to coordinate wider implementation and use of Humboldt Core.
* All development will be done and progress tracked in the Humboldt Core GitHub repository from TDWG GitHub organization.

## **Becoming Involved**

* Individuals having an interest in this work should contact the convener, and are invited to watch and contribute via the Humboldt Core GitHub repository.

## **History/Context**

Assessments on biodiversity change are strongly limited by available data on species distributions (Meyer et al. 2015, Proença, et al. 2016). From the many data types that inform species occurrence, incidental point records, such as those generated from museum specimens or citizen science contributions, had seen strong recent growth. This was facilitated by the development of the Darwin Core data standard (Wieczorek et al. 2012), which allows the integration and re-use of Occurrence data (e.g., Symbiota, Gries et al. 2014 and the Integrated Publishing Toolkit, Robertson et al. 2014). Due to their greater complexity and more multi-faceted communities of providers and users, other biodiversity data types from typically more formal monitoring efforts are, to date, lacking widely applicable standards (Guralnick et al. 2018). The Humboldt Core is an effort to capture standardized information about processes underpinning inventory work (Guralnick et al. 2018).

## **Summary**

This task group will explore concepts and methods of species inventories to fully integrate this type of observational data into existing data exchange schemas. To support proper capture of key information about inventories, current Humboldt Core terms (Guralnick et al. 2018) will be reviewed, revised as necessary, and integrated into one or more extensions to the Darwin Core if possible. If not, the work of the Task Group will be amended to pursue an alternative standard for this type of data within the TDWG corpus of standards. The outcome in either case will be to provide a framework and clear semantics for sharing and integrating biodiversity inventory data.

## **Resources**

* Humboldt Core resources:
  + Humboldt Core in Map of Life: <https://mol.org/humboldtcore/>
  + Humboldt Core current implementation: <https://github.com/MapofLife/humboldtcore>
  + Humboldt Core publication: Guralnick R, Walls R, 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>
* TDWG Documentation
  + TDWG Vocabulary Maintenance Specification (VMS): <https://github.com/tdwg/vocab/blob/master/vms/maintenance-specification.md>
  + TDWG Standards Documentation Standard (SDS): <https://github.com/tdwg/vocab/blob/master/sds/documentation-specification.md>
* Other resources and references:
  + Biological Collections Ontology (BCO): <http://www.obofoundry.org/ontology/bco.html>
  + GBIF Darwin Core Extension schema: <http://rs.gbif.org/schema/extension.xsd>
  + Gries C, Gilbert E, Franz N. (2014) Symbiota – A virtual platform for creating voucher-based biodiversity information communities. Biodiversity Data Journal 2: e1114. <https://doi.org/10.3897/BDJ.2.e1114>
  + Meyer C, Kreft H, Guralnick R, et al. (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6, 8221. <https://doi.org/10.1038/ncomms9221>
  + Proença V, Martin LJ, Pereira HM, et al. (2013) Global biodiversity monitoring: from data sources to essential biodiversity variables. Bio Conserv 213: 256-263. <https://doi.org/10.1016/j.biocon.2016.07.014>. Part B.
  + Robertson T, Döring M, Guralnick R, et al. (2014) The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9(8): e102623. <https://doi.org/10.1371/journal.pone.0102623>
  + Schmeller DS, Julliard R, Bellingham PJ, et al. (2015) Towards a global terrestrial species monitoring program. J Nat Conserv 25: 51-57. <https://doi.org/10.1016/j.jnc.2015.03.003>
  + Wieczorek J, Bloom D, Guralnick R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>