Want to Describe and Share Biodiversity Inventory and Monitoring Data? The Humboldt Extension for Ecological Inventories Can Help!

Yanina V Sica[‡], Wesley M Hochachka[§], Yi-Ming Gan[|], Kate Ingenloff[¶], Dmitry Schigel[¶], Robert D. Stevenson[#], Steven J. Baskauf[¤], Peter Brenton[«], Anahita J. N. Kazem[»], John Wieczorek[^]

‡ Yale University, New Haven, United States of America
§ Lab of Ornithology, Cornell University, Ithaca, New York, United States of America
| Royal Belgian Institute of Natural Sciences, Brussels, Belgium
¶ GBIF, Copenhagen, Denmark
# University of Massachusetts, Boston, United States of America
¤ Vanderbilt University Libraries, Nashville, United States of America
« Atlas of Living Australia, Canberra, Australia
» German Centre for Integrative Biodiversity Research, Leipzig, Germany
^ VertNet, Bariloche, Argentina

Corresponding author: Yanina V Sica (yanina.sica@yale.edu)

Abstract

Access to high-quality ecological data is critical to assessing and modeling biodiversity and its changes through space and time. The Darwin Core standard has proven to be immensely helpful in sharing species occurrence data (see Wieczorek et al. 2012, Global Biodiversity Information Facility, GBIF) and promoting biodiversity research following the FAIR principles of findability, accessibility, interoperability and reusability (Wilkinson et al. 2016). However, it is limited in its ability to fully accommodate inventory data (i.e., linked records of multiple taxa at a specific place and time). Information about the inventory processes is often either unreported or described in an unstructured manner, limiting its potential re-use for larger-scale analyses. Two key aspects that are not captured in a structured manner yet are: i) information about the species that were not detected during an inventory, and ii) ancillary information about sampling effort and completeness.

Non-detections (i.e., reported counts of zero) potentially enable more accurate and precise estimates of distribution, abundance, and changes in abundance. This becomes possible when variation in effort is used to estimate the likelihood that a non-detection represents a true absence of that taxon during the inventory. Currently, ecological inventory data, when shared at all, are typically discoverable through dataset catalogs (e.g., governmental data repositories) and supplementary materials to publications. With few exceptions, indexing of such data with the detail and structure needed has not been attempted at broad temporal and spatial scales, despite the potentially high value resulting from making inventory data more readily accessible.

To address these limitations in documenting inventory data using the Darwin Core, Guralnick et al. (2018) proposed the Humboldt Core. Subsequent discussions within the biodiversity standards community made it clear that greater integration could be achieved by creating an extension of the Darwin Core, rather than developing a new standard in isolation. Extension design work began in 2021 and progress has been reported by Brenton (2021) and Sica et al. (2022).

Over the last year the Humboldt Extension Task Group has sought advice from data providers and aggregators and updated its vocabulary terms. A challenging aspect has been creating terminology for the parent-child relationships (see Properties of Hierarchical Events) needed to describe surveys that may be as simple as a collection of checklists (one level of hierarchy) or as complex as species records from traps within plots along transects across habitats over multiple years (at least four levels of hierarchy). The Task Group has committed to completing a User Guide for the Humboldt Extension. Group members who contributed to the Darwin Core (Darwin Core Task Group 2009) and the Vocabulary Maintenance Specification (Vocabulary Maintenance Specification Task Group 2017) have provided valuable expertise on term refinement and process.

Through ratification of the Humboldt Extension as a Darwin Core Event extension, we expect to provide the community with a usable solution, tied to well-established data publication mechanisms, for sharing and using inventory data. This effort promises to overcome a key bottleneck in the sharing of critically important ecological data, enhancing data discoverability, interoperability and re-use while lowering reporting burden and data and metadata heterogeneity. Global data aggregation initiatives, such as GBIF, will benefit from this development as they develop their data models and the range of standards and extensions they support. We anticipate that the Humboldt Extension will be attractive both to data publishers and data users, by facilitating the representation and indexing of data in richer, more meaningful ways. Despite the data-intensive nature of fundamental ecological research and applied monitoring for management and policy, ecological data have remained as one of the FAIR data frontiers. We anticipate that the Humboldt Extension will address most data exchange needs of all professional communities involved.

Keywords

Darwin Core, species surveys, sampling, standards

Presenting author

Yanina V. Sica

Presented at

TDWG 2023

Acknowledgements

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Brenton P (2021) An implementation approach for the Humboldt Extension to Darwin Core. Biodiversity Information Science and Standards 5: 75350. https://doi.org/10.3897/biss.5.75350
- Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG). URL: http://www.tdwg.org/standards/450
- 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
- Sica Y, Ingenloff K, Yi-Ming GA, Kachian Z, Baskauf SJ, Wieczorek J, Zermoglio P, Stevenson R (2022) Application of Humboldt Extension to real-world cases. Biodiversity Information Science and Standards 6: 91502. https://doi.org/10.3897/biss.6.91502
- Vocabulary Maintenance Specification Task Group (2017) Vocabulary Maintenance Standard. Biodiversity Information Standards (TDWG). URL: http://www.tdwg.org/standards/642
- Wieczorek J, Bloom D, Guralnick R, Blum S, M. D, al. e (2012) Darwin Core: an evolving community-developed biodiversity data standard. PLoS One 7 (1): 29715. https://doi.org/10.1371/journal.pone.0029715
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18