BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e38411
doi: 10.3897/biss.3.38411

OPEN ACCESS

Conference Abstract

# Trait Data Integration from the Perspective of a Data Aggregator

Jennifer Hammock[‡], Katja S Schulz[‡]

[‡] Smithsonian Institution, Washington, United States of America

Corresponding author: Jennifer Hammock (hammockj@si.edu)

Received: 21 Jul 2019 | Published: 20 Aug 2019

Citation: Hammock J, Schulz KS (2019) Trait Data Integration from the Perspective of a Data Aggregator. Biodiversity Information Science and Standards 3: e38411. https://doi.org/10.3897/biss.3.38411

## Abstract

The Encyclopedia of Life currently hosts ~8M attribute records for ~400k taxa (March 2019, not including geographic categories, Fig. 1). Our aggregation priorities include Essential Biodiversity Variables (Kissling et al. 2018) and other global scale research data priorities. Our primary strategy remains partnership with specialist open data aggregators; we are also developing tools for the deployment of evolutionarily conserved attribute values that scale quickly for global taxonomic coverage, for instance: tissue mineralization type (aragonite, calcite, silica...); trophic guild in certain clades; sensory modalities.

To support the aggregation and integration of trait information, data sets should be well structured, properly annotated and free of licensing or contractual restrictions so that they are ‘findable, accessible, interoperable, and reusable’ for both humans and machines (FAIR principles; Wilkinson et al. 2016). To this end, we are improving the documentation of protocols for the transformation, curation, and analysis of EOL data, and associated scripts and software are made available to ensure reproducibility. Proper acknowledgement of contributors and tracking of credit through derived data products promote both open data sharing and the use of aggregated resources. By exposing unique identifiers for data products, people, and institutions, data providers and aggregators can stimulate the development of automated solutions for the creation of contribution metrics. Since different aspects of provenance will be significant depending on the intended data use, better standardization of contributor roles (e.g., author, compiler, publisher, funder) is needed, as well as more detailed attribution guidance for data users.

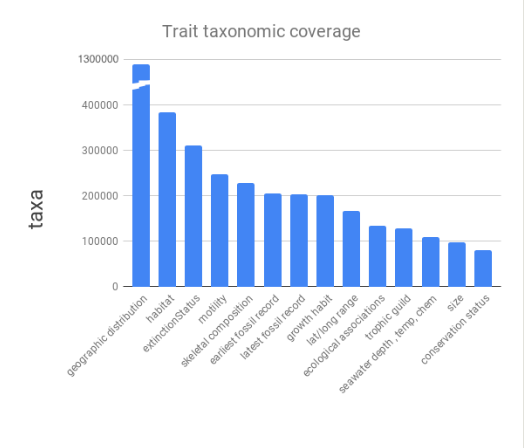

Figure 1.

Taxonomic coverage of trait categories in eol.org, March, 2019.

Global scale biodiversity data resources should resolve into a graph, linking taxa, specimens, occurrences, attributes, localities, and ecological interactions, as well as human agents, publications and institutions. Two key data categories for ensuring rich connectivity in the graph will be taxonomic and trait data. This graph can be supported by existing data hubs, if they share identifiers and/or create mappings between them, using standards and sharing practices developed by the biodiversity data community. Versioned archives of the combined graph could be published at intervals to appropriate open data repositories, and open source tools and training provided for researchers to access the combined graph of biodiversity knowledge from all sources. To achieve this, good communication among data hubs will be needed. We will need to share information about preferred vocabularies and identifier management practices, and collaborate on identifier mappings.

## Keywords

traits, data integration, graph data, identifiers

## Presenting author

Jennifer Hammock

## Presented at

Biodiversity_Next 2019

## Hosting institution

Smithsonian, National Museum of Natural History

## References

- Kissling WD, Walls R, Bowser A, Jones MO, Kattge J, Agosti D, Amengual J, Basset A, van Bodegom PM, Cornelissen JHC, Denny EG, Deudero S, Egloff W, Elmendorf SC, Alonso García E, Jones KD, Jones OR, Lavorel S, Lear D, Navarro LM, Pawar S, Pirzl R, Rüger N, Sal S, Salguero-Gómez R, Schigel D, Schulz K, Skidmore A, Guralnick RP (2018) Towards global data products of Essential Biodiversity Variables on species traits. Nature ecology & evolution 2 (10): 1531-1540. https://doi.org/10.1038/s41559-018-0667-3
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, 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, 't 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 S, 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, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3: 160018. https://doi.org/10.1038/sdata.2016.18