BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37749
doi: 10.3897/biss.3.37749

OPEN ACCESS

Conference Abstract

Distributed, but Global in Reach: Outline of a de-centralized paradigm for biodiversity data intelligence

Nico Franz‡, Edward Gilbert‡, Beckett W. Sterner‡

‡ Arizona State University, Tempe, United States of America

Corresponding author: Nico Franz (nico.franz@asu.edu)

Received: 28 Jun 2019 | Published: 02 Jul 2019

Citation: Franz N, Gilbert E, Sterner B (2019) Distributed, but Global in Reach: Outline of a de-centralized paradigm for biodiversity data intelligence. Biodiversity Information Science and Standards 3: e37749.
https://doi.org/10.3897/biss.3.37749

Abstract

We provide an overview and update on initiatives and approaches to add taxonomic data intelligence to distributed biodiversity knowledge networks. "Taxonomic intelligence" for biodiversity data is defined here as the ability to identify and renconcile source-contextualized taxonomic name-to-meaning relationships (Remsen 2016). We review the scientific opportunities, as well as information-technological and socio-economic pathways - both existing and envisioned - to embed de-centralized taxonomic data intelligence into the biodiversity data publication and knowledge intedgration processes.

We predict that the success of this project will ultimately rest on our ability to up-value the roles and recognition of systematic expertise and experts in large, aggregated data environments. We will argue that these environments will need to adhere to criteria for responsible data science and interests of coherent communities of practice (Wenger 2000, Stoyanovich et al. 2017). This means allowing for fair, accountable, and transparent representation and propagation of evolving systematic knowledge and enduring or newly apparent *conflict* in systematic perspective (Sterner and Franz 2017, Franz and Sterner 2018, Sterner et al. 2019).

We will demonstrate in principle and through concrete use cases, how to de-centralize systematic knowledge while maintaining *alignments* between congruent or concflicting taxonomic concept labels (Franz et al. 2016a, Franz et al. 2016b, Franz et al. 2019). The suggested approach uses custom-configured logic representation and reasoning methods, based on the Region Connection Calculus (RCC-5) alignment language. The approach offers syntactic consistency and semantic applicability or scalability across a wide range of biodiversity data products, ranging from occurrence records to phylogenomic trees. We will also illustrate how this kind of taxonomic data intelligence can be captured and propagated through existing or envisioned metadata conventions and standards (e.g., Senderov et al. 2018).

Having established an intellectual opportunity, as well as a technical solution pathway, we turn to the issue of developing an implementation and adoption strategy. Which biodiversity data environments are currently the most taxonomically intelligent, and why? How is this level of taxonomic data intelligence created, maintained, and propagated outward? How are taxonomic data intelligence services motivated or incentivized, both at the level of individuals and organizations? Which "concerned entities" within the greater biodiversity data publication enterprise are best positioned to promote such services? Are the most valuable lessons for biodiversity data science "hidden" in successful social media applications? What are good, feasible, incremental steps towards improving taxonomic data intelligence for a diversity of data publishers?

Keywords

alignment, biodiversity data, conflict, coordination, de-centralization, Region Connection Calculus, taxonomic concepts, systematics

Presenting author

Nico M. Franz

Presented at

Biodiversity_Next 2019

References

- Franz NM, Gilbert EE, Ludäscher B, Weakley AS (2016a) Controlling the taxonomic variable: Taxonomic concept resolution for a southeastern United States herbarium portal. Research Ideas and Outcomes 2 https://doi.org/10.3897/rio.2.e10610

- Franz NM, Pier NM, Reeder DM, Chen M, Yu S, Kianmajd P, Bowers S, Ludäscher B (2016b) Two influential primate classifications logically Aligned. Systematic Biology 65 (4): 561-582. https://doi.org/10.1093/sysbio/syw023
- Franz NM, Sterner BW (2018) To increase trust, change the social design behind aggregated biodiversity data. Database 2018 https://doi.org/10.1093/database/bax100
- Franz NM, Musher LJ, Brown JW, Yu S, Ludäscher B (2019) Verbalizing phylogenomic conflict: Representation of node congruence across competing reconstructions of the neoavian explosion. PLOS Computational Biology 15 (2). https://doi.org/10.1371/journal.pcbi.1006493
- Remsen D (2016) The use and limits of scientific names in biological informatics. ZooKeys 550: 207-223. https://doi.org/10.3897/zookeys.550.9546
- Senderov V, Simov K, Franz NM, Stoev P, Catapano T, Agosti D, Sautter G, Morris RA, Penev L (2018) OpenBiodiv-O: Ontology of the OpenBiodiv knowledge management system. Journal of Biomedical Semantics 9 (1). https://doi.org/10.1186/s13326-017-0174-5
- Sterner BW, Franz NM (2017) Taxonomy for humans or computers? Cognitive pragmatics for big data. Biological Theory 12 (2): 99-111. https://doi.org/10.1007/s13752-017-0259-5
- Sterner BW, Witteveen J, Franz NM (2019) Iternatives to realist consensus in bio-ontologies: Taxonomic classification as a basis for data discovery and integration. Preprint - ResearchGate 1-21. URL: https://www.researchgate.net/publication/331498583
- Stoyanovich J, Howe B, Abiteboul S, Miklau G, Sahuguet A, Weikum G (2017) Fides: Towards a platform for responsible data science. Proceedings of the 29th International Conference on Scientific and Statistical Database Management - SSDBM '17 https://doi.org/10.1145/3085504.3085530
- Wenger E (2000) Communities of practice and social learning systems. Organization 7 (2): 225-246. https://doi.org/10.1177/135050840072002