

Towards a Quality Assurance and Quality Control Mechanism for Species List Building

Olaf Bánki^{‡,§}, Donald Hobern^l, Markus Döring^{¶,§}, Geoffrey Donald Ower^{#,□}, Yuri Roskov^{#,□}, Diana Hernandez-Robles[«], Camila Andrea Plata[»], Peter H Schalk[§], Thomas M. Orrell[^]

‡ Naturalis Biodiversity Center, Leiden, Netherlands

§ Catalogue of Life / Species 2000, Weesp, Netherlands

l Catalogue of Life / Species 2000, Canberra, Australia

¶ Global Biodiversity Information Facility, Copenhagen, Denmark

University of Illinois, Champaign, Illinois, United States of America

□ Catalogue of Life / Species 2000, Champaign, Illinois, United States of America

« Catalogue of Life / Species 2000, Mexico City, Mexico

» Catalogue of Life / Species 2000, Bogotá, Colombia

^ National Museum of Natural History, Smithsonian Institution, Washington, DC, United States of America

Corresponding author: Olaf Bánki (olaf.banki@sp2000.org)

Abstract

[Catalogue of Life](#) (COL) brings together the efforts and contributions of taxonomists from around the world, and addresses the needs of researchers, policy-makers, environmental managers and the wider public for a consistent, up-to-date and authoritative listing of all the world's known species. The names that science gives to species are fundamental tools that allow us to refer to these units of biodiversity. Knowing the name for a species unlocks everything that has been learned about its biology, distribution and relevance to humans. Every day, taxonomists continue to publish new scientific names and refine our understanding of the world's species.

Over the past 10 years, game-changing progress has been made with the Catalogue of Life and its authoritative taxonomic name indexing services. Major steps were made towards content completion (over 2.1 million species out of all described 2.4 million are now documented and validated through COL (Bánki et al. 2023)). The COL data infrastructure and services were rebuilt in collaboration with the [Global Biodiversity Information Facility](#) (GBIF). This resulted in the GBIF-COL [ChecklistBank](#), an open global repository for publishing and sharing taxonomic and nomenclatural datasets (> 48,000 datasets, including over 160 Global Species Databases).

The data managing/editing process underlying the integration of data into a single checklist (index) could be more transparent with clearly documented protocols and procedures, and is today far too time consuming. In the ChecklistBank infrastructure, there are thousands of data sources at our disposal and we need a powerful, semi-automated, community-expertise based, quality assurance process to warrant the best possible product for users of taxonomic checklists.

COL aspires to document, with as much detail as possible, the current quality control and quality assurance processes as they relate to activities that occur at the following levels: 1) data providers, 2) pipeline and data conversions, 3) ChecklistBank infrastructure, and 4) checklist output. Ideally, this should happen throughout the consortium of partners that help the taxonomic community with mobilising and editing authoritative taxonomic data, including initiatives like [Atlas of Living Australia](#), [Catalogue of Life China](#), [EDIT Platform for Cybertaxonomy](#), [Integrated Taxonomic Information System](#), [Freshwater Animal Diversity Assessment](#), [TaxonWorks / Species Files](#), [World Flora Online / Rhakhis](#), [World Register of Marine Species / LifeWatch Aphia](#).

In recent years more attention has been given to the principles of creating an authoritative list of the world's species (Garnett et al. 2020). COL has developed and adopted new criteria for the inclusion of taxonomic data sources in the COL Checklist (Hobern et al. 2021). And ChecklistBank, certainly provides more avenues for (automated) quality checks and procedures.

But to guarantee the highest confidence in data, we are in need of a transparent and consistent data quality assurance and data quality control mechanism from the creation of taxonomic data, to data publishing, to the delivery into taxonomic data products and the global list of known species. We will present the current status of quality control and quality assurance processes in Catalogue of Life, and discuss what should constitute a quality control and quality assurance mechanism for species list building.

Keywords

Catalogue of Life, ChecklistBank, GBIF, taxonomy, nomenclature

Presenting author

Olaf Bánki

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Bánki O, Roskov Y, Döring M, Ower G, Hernández Robles DR, Plata Corredor CA, Stjernegaard Jeppesen T, Örn A, Vandepitte L, Hobern D, Schalk P, DeWalt RE, Keping M, Miller J, Orrell T, et al. (2023) Catalogue of Life Checklist (Annual Checklist 2023). Catalogue of Life. URL: <https://doi.org/10.48580/dfs>
- Garnett S, Christidis L, Conix S, Costello M, Zachos F, Bánki O, Bao Y, Barik S, Buckeridge J, Hobern D, Lien A, Montgomery N, Nikolaeva S, Pyle R, Thomson S, van Dijk PP, Whalen A, Zhang Z, Thiele K (2020) Principles for creating a single authoritative list of the world's species. PLOS Biology 18 (7). <https://doi.org/10.1371/journal.pbio.3000736>
- Hobern D, Barik S, Christidis L, T.Garnett S, Kirk P, Orrell T, Pape T, Pyle R, Thiele K, Zachos F, Bánki O (2021) Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution 21 (4): 677-690. <https://doi.org/10.1007/s13127-021-00516-w>