WF.ACTIAS: A workflow for a better integration of biodiversity data from diverse sources

Liliana Ballesteros Mejia, Rodolphe Rougerie, Sujeevan Ratnasingham

BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37139
doi: 10.3897/biss.3.37139

OPEN ACCESS

Conference Abstract

WF.ACTIAS: A workflow for a better integration of biodiversity data from diverse sources

Liliana Ballesteros Mejia[‡,§], Rodolphe Rougerie[‡], ACTIAS Consortium[|], Sujeevan Ratnasingham[¶]

‡ Muséum national d'Histoire Naturelle, Paris, France
§ CESAB, Centre de Synthèse et d'Analyse sur la Biodiversité, Montpellier, France
| Muséum National d'Histoire Naturelle, Paris, France
¶ University of Guelph, Guelph, Canada

Corresponding author: Liliana Ballesteros Mejia (ballesteros.liliana@gmail.com)

Received: 11 Jun 2019 | Published: 18 Jun 2019

Citation: Ballesteros Mejia L, Rougerie R, Consortium A, Ratnasingham S (2019) WF.ACTIAS: A workflow for a better integration of biodiversity data from diverse sources. Biodiversity Information Science and Standards 3: e37139. https://doi.org/10.3897/biss.3.37139

Abstract

Our knowledge of global biodiversity remains incomplete and beset by knowledge shortfalls affecting both the census of species (i.e. the Linnean shortfall) and our understanding of their distributions (i.e. the Wallacean shortfall; Hortal et al. 2015). While alarming rates of species extinction have been reported in most groups of organisms, our capacity to assess extinction threats is limited by these shortfalls and it has become imperative to optimize our use of existing information for the analyses of biodiversity data. There are two major challenges when integrating biodiversity data from heterogeneous sources to ultimately analyzing them:

1. The frequent disparity in taxon names used to refer to the same organisms;
2. Geographical inconsistencies on specimen information.

The first refers to disagreements about the taxon concepts attached to names alongside the different interpretations and applications (e.g. gender agreement in taxonomic names) of the existing nomenclatural rules that ensure universality and stability of scientific names. The development of new methods for species delineation, and in particular with the growing integration of genetic data in the practice of taxonomy (e.g. DNA barcoding; Ratnasingham and Hebert 2013), has increased our ability to discriminate closely related

species. This enhances the resolution level at which biodiversity is documented, described and analyzed(Goldstein and DeSalle 2011). One frequent outcome is the redefinition of species boundaries; either through merging (synonymy) or splitting of previously recognized species. In understudied groups such as insects, the resulting inflation of names, sometimes provisional, further defies the reconciliation of names used by different sources.

The second challenge refers to the completeness and accuracy of geographical information. Specimen records in biodiversity databases often lack GPS coordinates. Consequently we need to accurately infer the latitude and longitude from place names. Other frequent inaccuracies include erroneous georeferencing, imprecision and/or error in the location of a record (Soberón and Peterson 2004).

Integration of data on the basis of taxon names and their geographic information is a major challenge that either results in excluding a significant number of records, or in merging incompatible records, leading to erroneous outcomes. Therefore, we have developed WF.ACTIAS, a computational workflow that gathers data from several sources and provides the user with tools to make objective and reproducible decisions to assign records to a consensus species name, while detecting and correcting geographical inconsistencies. Its main objective is to automate a process that can integrate information about nomenclature, taxon concepts and geographical information to reconcile taxon names from different sources. Here, we present the WF.ACTIAS workflow in the context of the analysis of diversity in two sister families of moths – the Saturniidae and Sphingidae. Species boundaries in these insects have been thoroughly and comprehensively revisited through the integration of DNA barcodes and we are tackling the reconciliation of taxon names in ca. 282 000 records of which more than 77 000 have DNA barcodes. The outcome of this data integration is essential to study patterns of biodiversity and distributions and sets the ground to extend this process to other groups of organisms.

Keywords

Integrative taxonomy, biodiversity databases, insects, geographical information of specimens, biodiversity informatics

Presenting author

Sujeevan Ratnasingham

Presented at

Biodiversity_Next 2019

References

- Goldstein P, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. BioEssays 33 (2): 135-147. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/bies.201000036
- Hortal J, de Bello F, Diniz-Filho J, Lewinsohn T, Lobo J, Ladle R (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics 46 (1): 523-549. https://doi.org/10.1146/annurev-ecolsys-112414-054400
- Ratnasingham S, Hebert PN (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. PLoS ONE 8 (7). https://doi.org/10.1371/journal.pone.0066213
- Soberón J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359 (1444): 689-698. https://doi.org/10.1098/rstb.2003.1439