Ecological Indicators 90 (2018) 74–78

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Short Note

Nomenclature instability in species culturomic assessments: Why synonyms matter

Ricardo A. Correia[a,b,c,*], Ivan Jarić[d,e], Paul Jepson[c], Ana C.M. Malhado[b], José A. Alves[a,f], Richard J. Ladle[b,c]

[a] DBIO & CESAM-Centre for Environmental and Marine Studies, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
[b] Institute of Biological and Health Sciences, Federal University of Alagoas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, 57072-90 Maceió, AL, Brazil
[c] School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, United Kingdom
[d] Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, Na Sádkách 702/7, 370 05 České Budějovice, Czech Republic
[e] Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany
[f] University of Iceland, South Iceland Research Centre, Lindarbraut 4, IS-840 Laugarvatn, Iceland

A R T I C L E I N F O

Keywords:
Biodiversity
Data mining
Internet salience
Public perception
Scientific names
Species assessments
Taxonomy

A B S T R A C T

Culturomics is an emerging area of study that explores human culture through the quantitative analysis of large digital bodies of text. Culturomics shows great potential for the study of public perceptions and engagement with nature and biodiversity, and thus to contribute to the assessment and monitoring of major conservation goals (e.g. Aichi Target 1). In order to realize the full potential of culturomic approaches for conservation applications, researchers must develop solutions for existing methodological issues. For example, the use of scientific binomial names in species assessments has been recently proposed as a means to account for linguistic challenges associated with vernacular names, such as synonyms and homonyms. However, scientific names can also be affected by scientific synonyms arising from changes in species nomenclature. Here, we focus on a culturomic assessment of internet content and evaluate the importance of considering scientific name synonyms in such assessments. For this, we estimated how much omitting taxonomic synonyms affected webpage retrieval for bird species. Results indicate that failing to consider synonyms affected the number of webpages retrieved for over half of the species considered. In some cases, such omissions were severe (over 50% of total webpages omitted) and increased with the number of synonyms identified. We discuss the challenges posed by the dynamic nature of taxonomy in efforts to evaluate public interest in species using culturomic approaches and suggest that future studies should always strive to identify and account for any existing synonyms to minimize potential problems.

1. Introduction

The increasing application of social science methodologies is improving and extending our understanding of the human dimensions of conservation and natural resource policy and management (Bennett et al., 2017a,b). One particularly exciting new area of social science research applied to conservation is culturomics (Ladle et al., 2016), the exploration of human practices of engaging with nature through methods that focus on the quantitative analysis of large bodies of text (Michel et al., 2011). Culturomic techniques demonstrate great potential to contribute towards the assessment and monitoring of major conservation goals, such as improving public awareness of nature and its values (i.e. Aichi Target 1). Over the last five years, the use of culturomic techniques in conservation studies has increased considerably, especially as a way of gauging public interest in protecting nature (McCallum and Bury, 2013; Papworth et al., 2015; Proulx et al., 2014; Troumbis, 2017), as well as in more specific components such as natural areas (Correia et al., 2018b; Do et al., 2015) and species (Correia et al., 2016; Kim et al., 2014; Roll et al., 2016).

As in all emerging areas of research, culturomics researchers face numerous technical and methodological challenges. An important issue is how to deal with the presence of homonyms and synonyms which are common in biological *corpora* (Roll et al., in press). In the case of species assessments, an alternative is to use the (Latin) scientific name of a species rather than its vernacular name given the strong, consistent and culturally independent association between the representation of both name forms in digital corpora (Correia et al., 2017; Jaric et al., 2016). Using scientific names not only provides a potential solution for species with multiple vernacular names, it also allows researchers to carry out assessments for different language groups in a standardized manner

* Corresponding author at: DBIO & CESAM-Centre for Environmental and Marine Studies, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.
E-mail address: rahc85@gmail.com (R.A. Correia).

https://doi.org/10.1016/j.ecolind.2018.02.059
Received 8 November 2017; Received in revised form 18 February 2018; Accepted 28 February 2018
1470-160X/

(Correia et al., 2017).

However, the use of scientific names only partly solves the problem, since changes in species taxonomy or nomenclature generate taxonomic synonyms, which could also influence culturomic assessments. Taxonomic synonyms occur when more than one scientific name has been used to refer to a single species and fall into two major categories: orthographic or lexical synonyms and nomenclatural synonyms (Remsen, 2016). Orthographic synonyms represent variations in spelling of the same name originating from the variable application of grammatical rules (for example, differences in the Latin gender spelling such as *Aquila africana* and *A. africanus*), and typically originate from contemporary corrections of names where taxonomic rules were ignored or incorrectly applied. Nomenclatural synonyms are the result of taxonomic re-evaluations (e.g. splitting, clumping, or change of genus such as *Lutra felina* becoming *Lontra felina*). It should be noted that both forms of taxonomic synonyms also can cause problems for macroecological and biogeographical analysis, especially when datasets include species inventories collected over an extended time period (Tessarolo et al., 2017).

Here, we use Google's search engine to evaluate the importance of scientific synonyms for culturomic assessments of species representation on the internet. Specifically, we use a global list of bird species to: i) quantify the percentage of webpages omitted when taxonomic synonyms are not considered in web searches using scientific species names; ii) explore the relationship between the number of scientific name synonyms for a species and the number of omitted webpages; and iii) test how such omissions affect the relationship between scientific and vernacular name representation in internet corpora.

2. Material and methods

A global list of bird species containing 11121 individual species was obtained from BirdLife International (BirdLife International, 2016) – we considered this as our reference taxonomy. This list, recently developed using a new operational method for species delimitation (Tobias et al., 2010), considers the largest number of bird species and is currently recognized by a number of international policy and legislative bodies (Garnett and Christidis, 2017). For each species, we checked the existence of scientific name synonyms (hereafter, synonyms) in eight other taxonomic references (Table 1). Species were considered as the basic taxonomic unit for this study and synonyms were therefore only considered at the species level (i.e. subspecies synonyms were not considered). This assessment identified a total of 4913 individual synonyms. We considered all scientific names in the reference taxonomy plus all identified synonyms – a total of 16034 scientific names – for data collection.

Culturomic data collection followed the same methodological approach described in detail in Correia et al. (2017). In short, we used Google's Custom Search Engine API to carry out searches for each species using: i) the recognized vernacular name, ii) the recognized scientific name only, and iii) any identified synonyms. Searches for vernacular names and recognized scientific names were carried out during August 2017 by using quoted search strings (e.g. "Redpoll" OR "Common Redpoll"; "*Acanthis flammea*"), restricting results to exact matches of the search string, and the estimated number of webpages returned by each search was recorded. In the case of synonyms, searches were carried out during the same time period and using a similar method but excluding webpages mentioning the recognized name or other synonyms to avoid webpage double-counting (e.g. "*Carduelis flammea*"-"*Acanthis flammea*"). The complete dataset is freely available for consultation (Correia et al., 2018a).

For each species, we calculated the percentage of omitted webpages when considering only the recognized scientific name for web searches. This metric was calculated using the formula:

$$\frac{TOT-SCI}{TOT} \times 100$$

where TOT is the number of webpages returned using all species name synonyms and SCI is the number of webpages returned by a search using only the recognized species name. We also calculated the percentage of webpages originating from each individual synonym using the formula:

$$\frac{SYN}{TOT} \times 100$$

where SYN is the number of webpages returned by a search using only the synonym and TOT is the number of webpages returned using all species name synonyms. We then used histograms to analyse the distribution of omitted records across all species and the percentage of webpages contributed by each individual synonym. We evaluated the relationship between the proportion of omitted records and the number of synonyms recorded for each species with at least one recorded synonym using Generalized Linear Model (GLM) with binomial distribution and logit link following the formula:

$$logit\,(p) = \beta_0 + \beta_1 x$$

where p represents the proportion of omitted records, and x the number of synonyms recorded.

Finally, in line with previous assessments (Correia et al., 2017; Jaric et al., 2016), we analysed how the relationship between the log10 number of webpages retrieved using vernacular and scientific species names was affected by considering synonyms in web searches using Spearman's rank correlation. All analyses were carried out using R Software, and figures were produced in the same software using the *ggplot2* graphics package (Wickham, 2009). The analytical code used to produce the analysis is available in Appendix A.

Table 1
Taxonomic references used to identify scientific name synonyms for global bird species list. A comparison table of the taxonomic references considered in this study is available from the International Ornithological Congress website (http://www.worldbirdnames.org/).

Reference	Title
F. Gill, D. Donsker (2017)	IOC World Bird List (v 7.2)
E.C. Dickinson, J.V. Remsen Jr., L. Christidis (2013–2014)	The Howard & Moore Complete Checklist of the Birds of the World. 4th. Edition, Vol. 1, 2, Aves Press, Eastbourne, U.K
J.F. Clements, T.S. Schulenberg, M.J. Iliff, D. Roberson, T.A. Fredericks, B.L. Sullivan, C.L. Wood (2016)	The eBird/Clements checklist of birds of the world: v2016
J. del Hoyo, N.J. Collar, D.A. Christie, A. Elliott, L.D. C. Fishpool (2014)	HBW and BirdLife International Illustrated Checklist of the Birds of the World, Volume 1: Non-passerines. Lynx Editions
J.L. Peters (1931–1986)	Check-list of Birds of the World, vols. 1–16. Harvard University Press/Museum of Comparative Zoology
J.H. Boyd III (2017)	TiF checklist, Version 3.07
C.G. Sibley, B.L. Monroe (1993)	A Supplement to Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven, Connecticut
IUCN (2017)	The IUCN Red List of Threatened Species 2017-1