



# The species awareness index as a conservation culturomics metric for public biodiversity awareness

Joseph W. Millard <sup>1,2\*</sup> Richard D. Gregory <sup>1,3</sup> Kate E. Jones <sup>1</sup> and Robin Freeman<sup>2</sup>

<sup>1</sup>Department of Genetics, Evolution & Environment, University College London, Gower Street, London, WC1E 6BT, U.K.

<sup>2</sup>Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, U.K.

<sup>3</sup>RSPB Centre for Conservation Science, RSPB, The Lodge, Sandy, SG19 2DL, U.K.

**Abstract:** Although threats to global biodiversity are well known, slowing current rates of biodiversity loss remains a challenge. The Aichi targets set out 20 goals on which the international community should act to alleviate biodiversity decline, 1 of which (Target 1) aims to raise public awareness of the importance of biodiversity. Although conventional indicators for Target 1 are of low spatial and temporal coverage, conservation culturomics metrics show how biodiversity awareness can be quantified at the global scale. Following methods used for the Living Planet Index, we devised a species awareness index (SAI) to measure change in species awareness based on Wikipedia views. We calculated this index at the page level for 41,197 species listed by the International Union for Conservation of Nature (IUCN) across 10 Wikipedia languages and >2 billion views from 1 July 2015 to 30 March 2020. Bootstrapped indices for the page-level SAI showed that overall awareness of biodiversity increased marginally over time, although there were differences among taxonomic classes and languages. Among taxonomic classes, overall awareness increased fastest for reptiles and slowest for amphibians. Among languages, overall species awareness increased fastest for Japanese and slowest for Chinese and German users. Although awareness of species as a whole increased and was significantly higher for traded species, from January 2016 through January 2020, change in awareness appeared not to be strongly related to whether the species is traded or is a pollinator. As a data source for public biodiversity awareness, the SAI could be integrated into the Conservation International Biodiversity Engagement Indicator.

**Keywords:** Aichi Target 1, biodiversity, environmental awareness, indicator, online data, page views, Wikipedia

El Índice de Sensibilización de Especie como Medida de Culturomía de la Conservación para la Sensibilización Pública por la Biodiversidad

**Resumen:** Aunque las amenazas a la biodiversidad mundial son bien conocidas, reducir las tasas actuales de pérdida de la biodiversidad todavía es un desafío. Los objetivos de Aichi establecieron 20 metas para las cuales debe actuar la comunidad internacional para aliviar la declinación de la biodiversidad. Una de estas metas (Objetivo 1) busca sensibilizar al público sobre la importancia de la biodiversidad. Aunque los indicadores convencionales del Objetivo 1 tienen una baja cobertura espacial y temporal, las medidas de culturomía para la conservación muestran cómo la sensibilización por la biodiversidad puede cuantificarse a escala global. Seguimos los métodos utilizados para el Índice del Planeta Viviente para diseñar un índice de sensibilización de especie (ISE) para medir el cambio en la sensibilización por una especie con base en las vistas en Wikipedia. Calculamos este índice a nivel de página para 41,197 especies incluidas en las listas de la Unión Internacional para la Conservación de la Naturaleza (UICN) en diez diferentes idiomas en Wikipedia y más de 2 mil millones de vistas entre el 1 de julio de 2015 y el 30 de marzo de 2020. Los índices de arranque para el ISE a nivel de página mostraron que la sensibilización general por la biodiversidad incrementó ligeramente con el tiempo, aunque hubo diferencia entre

\*email joseph.millard.17@ucl.ac.uk

**Article impact statement:** Wikipedia represents a powerful data source for developing indicators of public biodiversity awareness. Paper submitted February 5, 2020; revised manuscript accepted October 22, 2020.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

las clasificaciones taxonómicas y los idiomas. Entre las clasificaciones taxonómicas, la sensibilización general incrementó más rápido para los reptiles y más lento para los anfibios. Entre los idiomas, la sensibilización general por especie incrementó más rápido para los usuarios del japonés y más lento para los usuarios del chino y el alemán. Aunque la sensibilización por las especies en su totalidad incrementó y fue significativamente más alta para las especies comercializadas, entre enero de 2016 y enero de 2020 el cambio en la sensibilización pareció no estar relacionado fuertemente con si la especie es un polinizador o es comercializada. Como fuente de información para la sensibilización pública por la biodiversidad, el ISE podría ser integrado dentro del Indicador de Participación Internacional para la Conservación de la Biodiversidad.

**Palabras Clave:** biodiversidad, conciencia ambiental, indicador, información en línea, Objetivo 1 de Aichi, vistas de página, Wikipedia

**摘要:** 尽管全球生物多样性面临的威胁已众所周知,但要减缓目前生物多样性丧失的速度仍然是一项挑战。“爱知目标”提出了 20 个目标来要求国际社会采取行动以缓解生物多样性下降,其中一个目标(目标 1)旨在提高公众对生物多样性重要性的认识。虽然目标 1 的传统指标的时空覆盖度较低,但保护文化组学指标展示了如何在全球范围内量化生物多样性意识。我们效仿地球生命力指数(Living Planet Index)的方法设计了一个物种认知指数,基于维基百科的浏览量来衡量人们对物种认知的变化。针对国际自然保护联盟(IUCN)列出的 41,197 个物种,我们在 10 种语言的维基百科上基于 2015 年 7 月 1 日到 2020 年 3 月 30 日期间超过 20 亿次页面浏览量计算了这个指数。基于网页的物种认知的引导指数显示,尽管在类群和语言之间存在差异,但人们对生物多样性的总体意识随时间推移都稍有提高。在不同类群中,人们对爬行动物的整体认识增长最快,而两栖动物的增长最慢。在不同语言中,日语使用者的总体物种认知增长最快,而中文和德语使用者的则增长最慢。尽管人们对物种的认识整体上有所提高,且对贸易物种的认识明显更多,但从 2016 年 1 月到 2020 年 1 月物种认知的变化似乎与该物种是否受贸易或是否为传粉者无关。物种认知指数作为公众生物多样性意识的数据来源,可以纳入国际生物多样性保护参与指标之中。【翻译:胡怡思;审校:聂永刚】

**关键词:** 爱知目标1, 维基百科, 指标, 网页浏览量, 在线数据, 生物多样性, 环境意识

## Introduction

Although threats to global biodiversity are well known, slowing current rates of biodiversity loss remains a challenge (Mace et al. 2018; IPBES 2019). Two problems are the requirement for transformational behavioral and economic change (IPBES 2019) and the difficulty in leveraging this change at a global level (IPBES 2019). The Strategic Plan for Biodiversity 2011–2020, underpinned by the Aichi Targets, represents an effort to guide these changes (UNEP CBD 2010). Specifically, the Aichi Targets set out 20 goals on which the international community should act to alleviate biodiversity decline (UNEP CBD 2010). Three of the Aichi Targets have sufficient and suitable indicators (6, 9, and 11), 4 have intermediately sufficient indicators (4, 7, 12, and 14), 10 have insufficient indicators (1, 5, 8, 10, 13, 16, 17–20), and 3 have none (Mcowen et al. 2016). Concerned with public awareness of biodiversity, Aichi Target 1 states that by 2020, the public should be aware of the value of biodiversity. Conventional indicators for Target 1 (i.e., the biodiversity barometer [UEBT 2019]) are of low spatial and temporal coverage (Leadley 2013; Mcowen et al. 2016) and do not incorporate awareness of biodiversity itself (i.e., species). Without robust metrics capturing evidence toward achieving target 1, determining whether this target has been met will be hard.

Conservation culturomics has emerged as a field concerned with digitized data and human–nature interactions (Ladle et al. 2016; Sherren et al. 2017). Quantifying public awareness of biodiversity is an area of active interest. Using data sources, such as Twitter, Facebook, Flickr,

Wikipedia, and Google Trends, a number of researchers have shown how online data can be used to improve understanding of how the public perceives biodiversity and environmentalism (e.g., McCallum & Bury 2013; Roberge 2014; Papworth et al. 2015). More recently, researchers have explored how online data sources can be combined to build a single indicator of biodiversity awareness. For example, Cooper et al. (2019) examined frequencies of biodiversity keywords across social media, online newspapers, and internet searches, reasoning that relative frequencies reflect public awareness of conservation issues. A significant step forward in applying culturomic approaches to the development of indicators, Cooper et al. (2019) provided a global framework for future research. Given their focus on conservation issues, a potential improvement could be to incorporate changing awareness of biodiversity itself.

Wikipedia page views represent a powerful data source for quantifying change in public awareness of biodiversity. Page views have been used to quantify public interest in reptiles (Roll et al. 2016) and species phenology (Mittermeier et al. 2019). In the context of awareness, Wikipedia is valuable in that pages are linked explicitly to biodiversity across scales. Pages on Wikipedia exist for taxa at multiple taxonomic levels, red-list statuses, and ecological systems and have an unambiguous link between the taxon and page identity (Mittermeier et al. 2019). Wikipedia can reveal changes in public awareness in response to natural history documentaries, demonstrating that the data source could be informative of long-term changes in awareness (Fernández-Bellón & Kane 2020). Moreover, because species characteristics

provide a mechanistic link to ecosystem services, change in awareness for a particular species on Wikipedia could be used as a proxy for awareness of its contribution. For example, increasing awareness for species that contribute significantly to pollination or trade could indicate greater public awareness of biodiversity importance. For pollination specifically, such changes in awareness are particularly important, given the global economic importance and reported declines of animal pollinators (e.g., Hallmann et al. 2017; IPBES 2016; Powney et al. 2019). Although using Wikipedia for quantifying awareness is not without its limitations and caveats (see Discussion), it provides the basis for a useful new indicator.

An awareness metric based on Wikipedia page views could be thought of as analogous to the Living Planet Index (LPI). The LPI represents an aggregation of vertebrate population trends (Loh et al. 2005; Collen 2009; McRae et al. 2017), showing the average rate of change for multiple species populations. Treating species page views as a population size, the LPI method could be similarly applied to Wikipedia to derive a rate of change for species awareness. Multiple researchers have used page views or search trends to infer change in awareness of specific species (e.g., Fink et al. 2020; Fukano et al. 2020; Lenda et al. 2020), but as far as we know, no one has calculated such an aggregated index for overall awareness. We devised and evaluated an approach based on the frequency of Wikipedia views for species listed on the International Union for Conservation of Nature (IUCN) Red List (hereafter IUCN species) that we call the species awareness index (SAI). We then explored variation in this metric, aiming to assess whether awareness of biodiversity has changed. Specifically, we explored the overall SAI for 41,197 IUCN species pooled and 6 distinct taxonomic classes—including the core pollinating groups (insects, birds, and mammals) and the most heavily traded vertebrates (reptiles, mammals, birds, amphibians, and ray-finned fishes)—and the SAI for each taxonomic class in each of the top 10 languages (by active user) on Wikipedia (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish). We then modeled rate of change in the page-level SAI as a function of taxonomic class, Wikipedia language, trade contribution, and pollination contribution with a pollinator data set derived from the academic literature through named-entity recognition. We considered the limitations of the SAI and how the SAI might be combined with other approaches for a more holistic understanding of changing biodiversity awareness.

## Methods

### Wikipedia Data

We used the Wikipedia page view API (application programming interface), and software written in Python, to

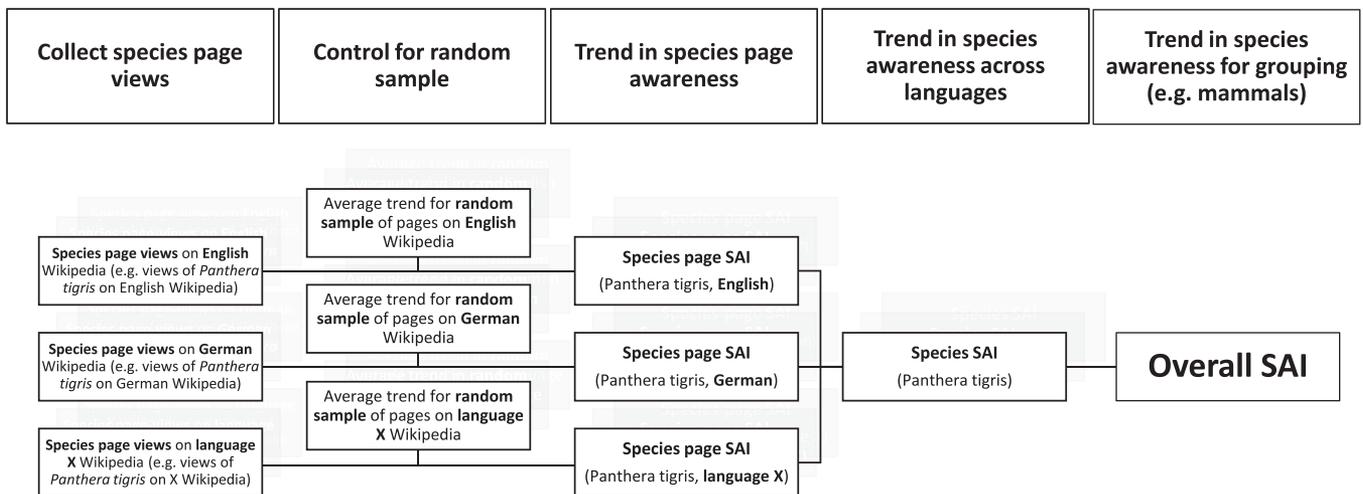
download daily user views of IUCN species for the period 1 July 2015–31 March 2020 (downloaded on the 16–21 April 2020). We downloaded views for all IUCN species with Wikipedia pages in the taxonomic group reptiles, ray-finned fishes (Actinopterygii), mammals, birds, insects, and amphibians from 10 Wikipedia language projects (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish). We retrieved our list of IUCN species on Wikipedia from OneZoom (Rosindell & Wong 2020; Wong & Rosindell 2020), which used the Wikipedia API to map between a species' scientific name, IUCN identifier (ID), Wikidata Q identifier, and the main Wikipedia page name for each species in each language. Downloading views from only the main page name of each species excludes redirect views and thus controls for potential variation caused by the URL used to reach a page. Each IUCN ID is unique to a species on the IUCN database (IUCN 2020), whereas each Wikidata Q identifier is unique to 1 species for all the Wikipedia languages in which that species appears. We refer to each individual Wikipedia page for a particular species in a given language as a *species page* to distinguish these pages from our use of *species* to refer to a particular species among languages.

For each species page, we retrieved only user views (i.e., views for which the visitor to that page was recorded as human, excluding automated views from bots). As in Mittermeier et al. (2019), we were not able to retrieve views from before 1 July 2015 because views from before this date are not archived by Wikipedia at the page-view API. For each species page returned, we calculated the daily average views for each month, and then kept only those species pages for which the series was represented for all months (see Appendix S6 for number of complete series). We used daily average views rather than total views because the Wikipedia page-view API does not always return views for all days in a given month.

To account for the overall change in Wikipedia's popularity and use, we also downloaded the daily user views for a random set of 11,000 pages in each language with the Wikipedia Random API, which requests random pages. We then aggregated these views in the same manner as the daily average species views and again kept only pages represented across the whole time series. From this random set of views, we then removed any page also appearing in the set of species pages for that language. We initially sampled 11,000 pages to maximize the number of remaining pages after removing incomplete series and species pages.

### Pollinator and Wildlife Trade Data Sets

To explore how species awareness varied with pollination contribution, we built a list of animal pollinators by combining text analysis and manual inspection of



**Figure 1.** A schematic of how the species page species awareness index (SAI) (i.e., random adjusted trend for a given species in a given language), species SAI (average of species page SAIs for a single species across languages), and overall SAI (group of bootstrapped species SAIs) were derived using Wikipedia views. Gray shadow rectangles represent multiple potential species, which are bootstrapped to calculate an overall SAI. For each species page, the trend in page views is adjusted for the average change in a random set of nonspecies pages (“control for random sample”).

the pollination literature (see Millard et al. [2020] and Appendix S1 for detailed methods). We also used the list of traded vertebrate species released in Scheffers et al. (2019) and the Food and Agriculture Organization (FAO) fisheries statistics (FAO 2020) to compile a data set of traded mammals, birds, squamate reptiles, and ray-finned fish. We then retrieved the Wikidata Q ID for each of these traded species with the Wikipedia API, which we merged onto each species page. We considered that all pollinator species make a pollination contribution and species listed in either Scheffers et al. (2019) or the FAO statistics make a trade contribution.

### Calculating Absolute Awareness of Biodiversity

Before calculating the SAI, we briefly explored absolute awareness of biodiversity among taxonomic classes, pollination contribution, and trade contribution. We defined *absolute awareness* as the total views for a species page on Wikipedia from 1 July 2015 to 31 March 2020. We combined the total views for each species page with the taxonomic class, trade contribution, and pollination contribution of that species and then built 2 generalized linear mixed-effects models: a model of  $\log_{10}$  total article views as a function of taxonomic class, trade contribution (yes or no), the interaction of class and trade, and a random effect for language and a model of  $\log_{10}$  total article views as a function of taxonomic class, pollination contribution (yes or no), the interaction of class and pollination, and a random effect for language. Rather than attempting to find the most parsimonious model, we determined full model predicted values, with AIC values

for these and a set of candidate null models described in Appendices S28 and S29.

### Deriving the SAI

The SAI is a new measurement of change in species awareness calculated at the species page level from the rate of change in daily average Wikipedia views per month. Because the SAI measures the rate of change in views within a species page, species are weighted equally irrespective of their popularity, meaning highly viewed species do not dominate the SAI. Hereafter, we use *SAI* or *species awareness index* to refer to the overall change in awareness for a given species page, species, or group of species on Wikipedia. Specifically, we used *species page SAI* to refer to rate of change at the page level, *species SAI* to refer to the average of all species page SAIs for a unique species among languages, and *overall SAI* to refer to a bootstrapped group of species SAIs (Fig. 1). We used *average monthly rate of change in the species page SAI* to refer to the average rate of change for a single species page across a given period. All of the above are distinct from absolute interest in a given species or group of species (i.e., the total Wikipedia views over the whole time series).

To construct the species page SAI, we used the R package *rlpi* to calculate an index of change over time for each species in 6 taxonomic groups (amphibians, birds, insects, mammals, ray-finned fish, and reptiles) on 10 Wikipedia languages (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish). The *rlpi* R package applies a generalized

additive model to smooth the daily average species page view trends ( $k = N/2$  is degrees of freedom), following Collen (2009). In *rlpi*, these smoothed values are then used to calculate a rate of change in views for a species page article:

$$\lambda_{st} = \log_{10} \left( \frac{N_t}{N_{(t-1)}} \right), \quad (1)$$

where  $\lambda_s$  is the rate of change in a species page,  $N$  is the smoothed number of daily average page views per month, and  $t$  is month.

To account for the overall change in popularity of Wikipedia itself over the same period, we adjusted the rate of change for each species page based on the rate of change in a random set of complete series Wikipedia pages (number of complete series in Appendices S6 & S9). For each species page, this adjustment was made with a random set of pages in the Wikipedia language of that species page. For example, the Wikipedia page for *Panthera tigris* in the English language would be adjusted for a set of random pages in the English Wikipedia, whereas the page for *P. tigris* in French would be adjusted for a set of random pages in the French Wikipedia. To do so, we first calculated the rate of change for each random page in each language with *rlpi*, as in species pages. We then used a bootstrap resampling approach to calculate the average rate of change for all random pages in a given language at each time step. The average rate of change in the random pages ( $\lambda_{rt}$ ) was calculated by bootstrapping the monthly rates of change 1000 times and then extracting the bootstrapped mean. At each time step, we then adjusted the species page rate of change by subtracting the monthly bootstrap estimated random rate of change ( $\lambda_{rt}$ ):

$$\lambda_{at} = \lambda_{st} - \lambda_{rt}, \quad (2)$$

where  $\lambda$  is rate of change,  $t$  is month,  $r$  is bootstrapped random trend for a given language,  $s$  is species page trend for that same language, and  $a$  is adjusted species trend.

For each species page, the SAI is then

$$I_{at} = I_{a(t-1)} * 10^{\lambda_{at}}, \quad (3)$$

where  $I_{a0} = 1$  and  $I_{at}$  is the species page SAI at time  $t$ .

To account for differences in the tortuosity of trends among Wikipedia languages (see Appendix S12), we also smoothed the species page SAI in each Wikipedia language with a loess regression (span = 0.3) before transforming the smoothed species page SAI back into a rate of change.

After smoothing the species page SAI as above, we calculated a species SAI for each species (across languages) by averaging rates of change at each time step across all languages. For example, the species *P. tigris* has the unique Wikidata ID Q19939, meaning the average rate of change in SAI for all species pages (irrespective of lan-

guage) identified as Q19939 provided the overall rate of change for the species *P. tigris*.

We then calculated an overall SAI combining all species across 10 Wikipedia languages by averaging rates of change across all species SAIs. Bootstrap confidence intervals were calculated by taking the 2.5th and 97.5th percentiles of 1000 bootstrapped indices at each time step. To check the extent to which single languages influenced the overall SAI, we jackknifed the overall SAI for language and removed languages with a marked effect on the overall trend (Appendix S11).

Using the same approach as above, we calculated an overall SAI for each taxonomic class for all languages combined and for each taxonomic class in each language. For each taxonomic class, we again averaged the loess smoothed rate of change in species page SAI among languages and then bootstrapped the species rate of change in SAI at each time step for each taxonomic class, as above. To check the extent to which single languages influence class-level trends, we again jackknifed the overall SAI for language and removed languages with a marked effect on the overall trend (Appendix S13). To calculate an overall SAI in each taxonomic class in each language, we bootstrapped the rate of change in species page SAI for the set of species pages in a given class and language combination.

### Modeling Average Monthly Rate of Change in the SAI

After calculating the SAI for all species pages on Wikipedia, we calculated an average monthly rate of change in each smoothed species page SAI for January 2016–January 2020. This average monthly rate of change was calculated across complete yearly periods to control for the effect of seasonality. To robustly explore whether change in awareness differed for various groups, we constructed 1 linear model and 2 mixed-effects linear models in which we fitted average monthly rate of change in species page SAI. The linear model was for average rate of change in species page SAI as a function of taxonomic class, language, and their interaction. One mixed-effects model was for average rate of change in species page SAI as a function of taxonomic class, pollination contribution (yes or no), their interaction, and a random effect for language and the other was for average rate of change in species page SAI as a function of taxonomic class, traded status (yes or no), and a random effect for language. Rather than attempting to find the most parsimonious model, we determined full model predicted values and AIC values for these and a set of candidate null models (Appendices S30, S31, & S32). With the exception of API requests made in Python, all analyses and data processing were carried out in R 4.0.3 (R Core Team 2020).

## Results

### Wikipedia View Data Set

Before removing incomplete series, our initial Wikipedia data set included approximately 2.23 billion page views for IUCN species across the 10 Wikipedia languages. These views were represented across 41,197 IUCN species, over 1735 days between 1 July 2015 and 31 March 2020. Views for each language varied from approximately 24.92 million views in the Arabic Wikipedia to approximately 1.08 billion views in the English Wikipedia (Appendix S4). For all languages, unique species number was highest for ray-finned fishes at 13,571 and lowest for insects at 2743 (Appendix S4; full language breakdown in Appendix S6). The series subset for only pages represented across all months showed the proportion of complete series was lowest in the Arabic Wikipedia, specifically the ray-finned fishes (approximately 35%) and the reptiles (approximately 38%). Most taxonomic classes for most languages had complete series in at least 80% of the species in that grouping (Appendix S6).

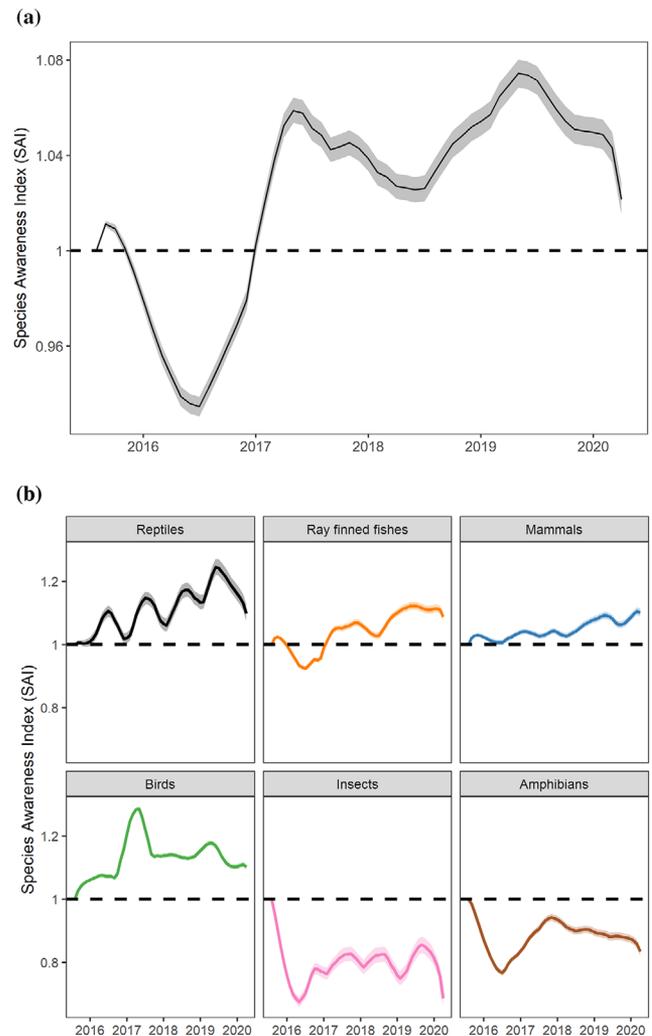
After removing pages also present in the species set, our set of random views consisted of approximately 2.82 billion views across 113,622 random pages (Appendix S8), again for the same 1735 days. The total number of random views was highest for the English Wikipedia at approximately 629.85 million views and lowest in the Arabic Wikipedia at approximately 87.94 million views (Appendix S8). The subset for only random pages represented for all months showed total random pages varied from 3486 in the Arabic Wikipedia to 9174 in the Japanese Wikipedia (Appendix S9).

### Absolute Awareness of Biodiversity

Among taxonomic classes, reptiles had consistently higher absolute awareness, appearing in the top 2 classes for 7 of 10 languages (Appendix S16). In contrast, amphibians had consistently lower awareness, appearing in the bottom 2 classes for 8 of 10 languages. Some languages appeared to have uniquely high absolute awareness for specific classes. For example, the ray-finned fishes had the highest absolute awareness in the Japanese Wikipedia (Appendix S16). Across all languages, absolute awareness (total views) was significantly higher in traded species ( $F = 15206.44$ ,  $p < 0.001$ ) (Appendices S5 & S24), but not significantly different in pollinating species ( $F = 0.3869$ ,  $p = 0.5339$ ) (Appendix S23).

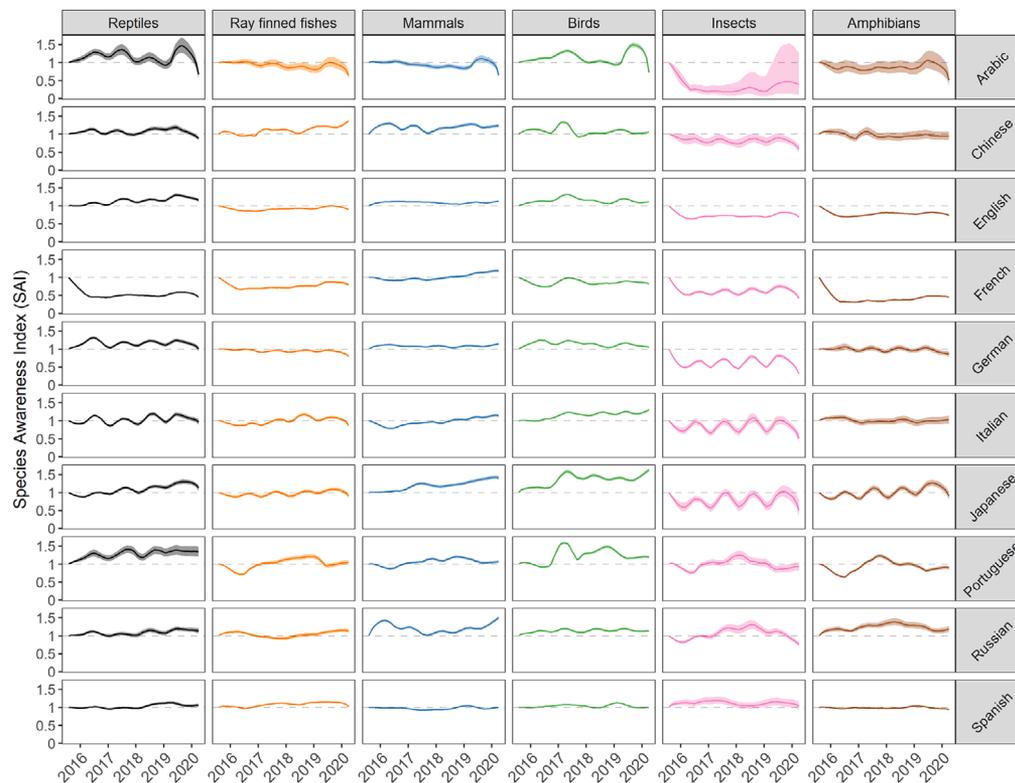
### The SAI

The overall SAI for all taxa and languages was markedly affected by the inclusion of the French Wikipedia (Appendix S11), so we excluded it from analyses of aggreg-



**Figure 2.** The species awareness index (SAI) for reptiles, ray-finned fishes, mammals, birds, insects, and amphibians on the Wikipedia languages Arabic, Chinese, English, German, Italian, Japanese, Portuguese, Russian, and Spanish for July 2015–March 2020 (a) overall and (b) by class (lines, mean of bootstrapped indices at each monthly time step; shading, 2.5th and 97.5th percentiles). Taxonomic class panels are ordered by magnitude of overall increase in each taxonomic class. French Wikipedia is not included given its marked influence on the aggregated SAI (see Appendices S11 & S13).

gated change at the overall level. With the exclusion of the French Wikipedia, the overall SAI increased over the whole series from July 2015 to March 2020 but declined markedly in January 2016 to June 2016 and June 2017 to May 2018 (Fig. 2b). The overall increase in the SAI was largely robust to variable baselines (Appendix S22), although average rate of change was marginally negative from mid-2017 and highly negative from mid-2019.



**Figure 3.** The species awareness index (SAI) for 6 taxonomic classes across 10 Wikipedia languages for July 2015–March 2020 (lines, mean of bootstrapped indices at each monthly time step; shading, 2.5th and 97.5th percentiles). Taxonomic class panels are ordered by the magnitude of overall increase in each taxonomic class and for language alphabetically.

At the level of taxonomic class, jackknifing trends by language again showed that the French Wikipedia was markedly affecting the overall trend (Appendix S13). With the exclusion of the French Wikipedia, from July 2015 to March 2020, awareness of the reptiles, ray-finned fishes, mammals, and birds appeared to increase, whereas awareness of amphibians and insects appeared to decrease (Fig. 2a). Birds experienced a peak in early 2017 (Fig. 2b), driven by an increase across multiple languages (Fig. 3). Awareness of mammals consistently and steadily increased, particularly in the Japanese Wikipedia (Fig. 3). Awareness of amphibians and insects both dropped considerably from the start of the series to mid-2016 before increasing, the cause of which is unclear. The trend for reptiles and insects was highly seasonal for multiple languages, peaking in July–August of each year; the notable exception to this was the English language for insects (Fig. 3).

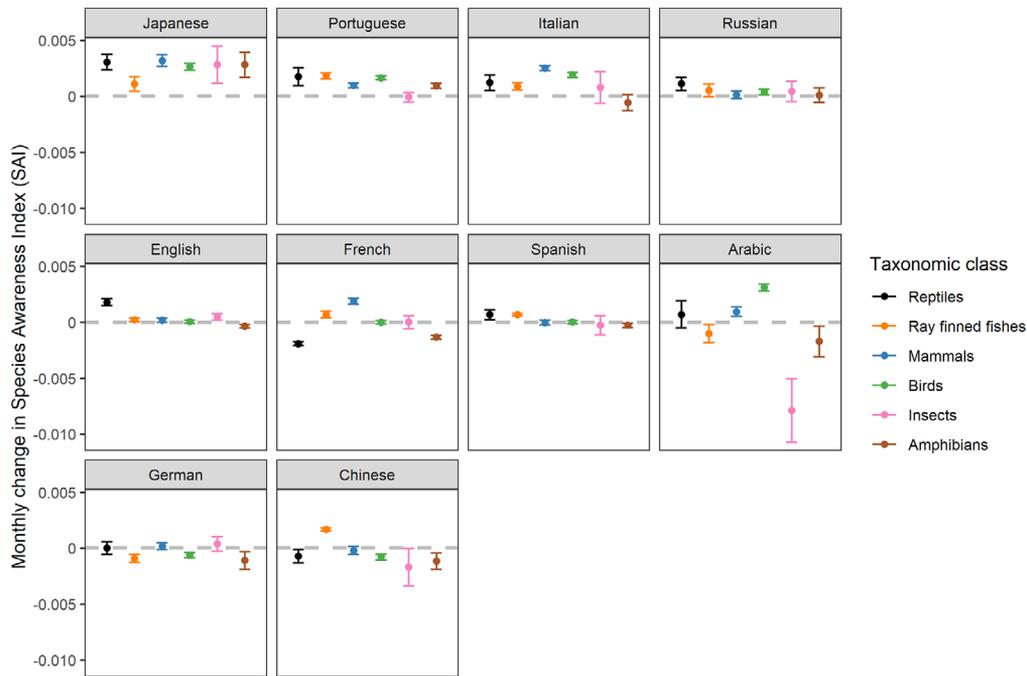
#### Modeling Average Monthly Rate of Change in Species Page SAI

Average monthly rate of change in species page SAI for January 2016–January 2020 differed significantly for taxonomic class, language, and their interaction (Fig. 4 & Appendix S27). At the level of taxonomic class, aware-

ness of reptiles and ray-finned fishes increased the fastest, and it increased or declined slowly for insects and amphibians (with the exception of the Japanese Wikipedia). Among languages, rate of change in species page SAI was highest in the Japanese and Portuguese Wikipedias and lowest in the German and Chinese Wikipedias. Although absolute interest was significantly greater in traded species (Appendices S5 & S24), from January 2016 to January 2020, average monthly rate of change in the species page SAI appeared not to be related to either trade contribution (Appendices S26 & S19) or pollination contribution (Appendices S25 & S19).

#### Discussion

The SAI is an index of change in public awareness of biodiversity derived from views of individual Wikipedia species pages. It enables investigation based on a variety of variables, such as taxonomy, language, geographic distribution, and ecosystem service provision. We found that awareness of biodiversity overall increased marginally; increases were highest for reptiles and ray-finned fish. Although biodiversity awareness increased overall, awareness of some groups (i.e., the amphibians



**Figure 4.** Average monthly rate of change for the species page species awareness index (SAI) for 6 taxonomic classes across 10 Wikipedia languages (error bars, predicted values of a linear model, fitting average monthly change in the species page SAI as a function of taxonomic class, Wikipedia language, and their interaction). Fitted values are from the linear model with the R function `predict` (points), and 95% CIs are from the fitted values  $\pm 1.96$  multiplied by the SE.

and insects) decreased or increased only marginally. Furthermore, our results suggest that change in awareness of biodiversity is likely not related to a species' trade contribution or pollination contribution. As an indicator of biodiversity awareness, a Wikipedia-derived metric, such as the SAI, represents a useful additional data source, given its explicit and unambiguous link to biodiversity itself at multiple scales (i.e., species, family, and class).

The link between culture and perceived biodiversity value or awareness is widely recognized (Daniel et al. 2012; Roll et al. 2016; Cooper et al. 2019; Ladle et al. 2019), but for a culturomics metric, such as the SAI, the drivers of change are complex. Overall trends capture many different drivers of awareness, making it difficult to isolate the causes for a given increase or decrease. The Chinese Wikipedia, for example, shows a consistent decrease in awareness for 5 taxonomic classes, but a consistent increase for ray-finned fish. We hypothesized that this increase for ray-finned fish may be driven by increasing fish consumption; seafood demand in China has increased significantly in recent years (FAO 2020). However, in a brief additional analysis, we found no significant difference between the rate of change for traded and nontraded ray-finned fish in the Chinese Wikipedia (Appendix S18). This indicates that the greater rate of change for the Chinese ray-finned fish may not be driven

by consumption alone. The Japanese Wikipedia is also of note, given its consistent increase in awareness across all 6 taxonomic classes. This awareness increase was concordant with the results of conventional surveys in which Japan had among the largest percentage point increases for familiarity with the term *biodiversity* (UEBT 2019). However, it is unclear what may be driving this change in awareness. Counterfactual scenario modeling could improve the understanding of such relationships, as has been demonstrated in a number of recent conservation culturomics studies (Acerbi et al. 2020; Fernández-Bellón & Kane 2020; Veríssimo et al. 2020).

Although drivers of overall change in the SAI are complex, it is conspicuous that absolute awareness and change in awareness did not differ between pollinating and nonpollinating animals. Because traded species have high absolute awareness relative to nontraded species, biodiversity awareness likely does relate to its value. But for pollination contribution, this relationship appears to be weak or nonexistent. For 3 main reasons, this may be the case. First, the impact pollination contribution has on biodiversity awareness will be highly taxa dependent; groups more strongly associated with pollination will have a larger increase in awareness. Second, unlike a species traded for direct consumption (i.e., food), pollinators make an indirect contribution to people, making their benefit less intuitive. Third, the nature of a

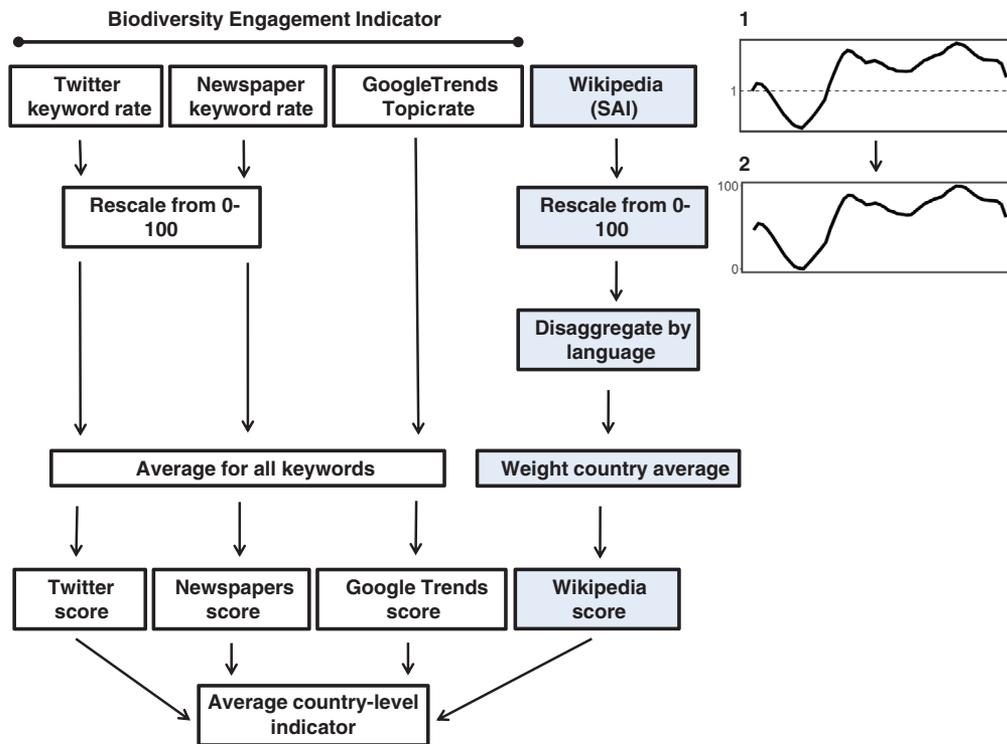


Figure 5. A method through which the species awareness index (SAI) could be incorporated in the biodiversity engagement indicator (BEI) for a single indicator of country-level biodiversity awareness (adapted from Cooper et al. [2019]) (white squares, methodological pathway for the original BEI; blue squares, potential additional pathway for combining the BEI with the SAI). Graph 1 shows the overall SAI scaled starting at 1. Graph 2 shows the overall SAI rescaled from 0 to 100, consistent with the BEI.

pollinator is that it is often not deliberately sought by those who benefit from its contribution. Given its central role in the value of biodiversity, more work is required to understand the contribution ecosystem service provision makes to public biodiversity awareness. Because pollinators have been so well publicized (Smith & Saunders 2016), one would expect that if awareness is not relatively high for pollinating species, it likely will not be for other service-providing species.

The SAI and Conservation International Biodiversity Engagement Indicator (BEI) have value as independent metrics, but as in Cooper et al. (2019), we emphasize the importance of combining multiple online platforms for inferring public biodiversity awareness. Particularly because the SAI provides an explicit link to biodiversity itself, its inclusion could provide a more holistic understanding as to how biodiversity awareness is changing. Combining the SAI with the BEI presents 2 core challenges. First, the BEI and SAI have different units of measure. The BEI is scaled on a 0–100 scale in a manner analogous to Google Trends. The SAI, in contrast, is scaled relative to a benchmark index of 1, in an approach inspired by the LPI. Second, there are problems of geographic scale in combining the BEI and SAI. Namely, the

BEI is aggregated at the country level, whereas the SAI cuts across countries at the language level.

Given differences in units and geographic scale, combining a Wikipedia metric and the BEI is not simple. One potential solution could be to rescale the SAI on a 0–100 scale, disaggregate by language, and then calculate a weighted average among languages to reflect the proportion of users for a given country (Fig. 5). The data for calculating such a weighting of language-level trends are provided by Wikipedia in a format amenable to web scraping (Wikimedia Traffic Analysis Report, 2018). Such an approach would solve the problems of differing units and geographic scales, transforming the SAI into a national metric amenable to averaging with Twitter, newspaper, and Google Trends scores (Fig. 5).

Despite providing a novel approach, online-derived metrics for biodiversity awareness are subject to limitations (Appendix S3). Primarily, online metrics are proxies rather than direct measures of awareness. For the SAI, one can show whether views increased or decreased on average for a given grouping, from which exposure to species-related information can be inferred. However, one cannot tell why a given page was visited or whether information related to that page was retained. On-site

Google Analytics can be used to return a suite of metrics that intimate reasons for a given visit (Soriano-Redondo et al. 2017), such as the site used to reach a page, but for Wikipedia, these data are not publicly available. Text mining could help quantify the type of information users are exposed to on Wikipedia. For example, calculating text similarity for each species page to a reference text on pollination would provide an indication of the pollination salience of a given species page. Similar approaches have been applied in the context of climate change and invasive species; rate of threat-related terms was used as an indicator of threat salience (Jarić et al. 2020).

As global internet penetration increases and biodiversity continues to decline, digital metrics for public biodiversity awareness will become more informative and important. The SAI showed that overall awareness of biodiversity appears to be increasing marginally, although the increase was inconsistent among taxonomic groups and languages. We also found that such increases appeared not to be related to trade or pollination contribution of species. We believe that combining the SAI with the BEI can provide a more holistic understanding of public biodiversity awareness in the digital realm.

## Acknowledgments

We thank Y. Wong and J. Rosindell for providing taxonomic mapping data and T. Newbold, R. Gibb, the CBER Journal Club and Data Club, and D. Redding for discussions and comments on earlier drafts of the manuscript. We also thank R. Correia and 3 anonymous reviewers for their valuable comments during review.

## Supporting Information

Additional information is available online in the Supporting Information section at the end of the online article. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

## Literature Cited

- Acerbi A, Kerhoas D, Webber AD, McCabe G, Mittermeier RA, Schwitzer C. 2020. The impact of the “World’s 25 Most Endangered Primates” list on scientific publications and media. *Journal for Nature Conservation* **54**:125794.
- Collen B. 2009. Monitoring change in vertebrate abundance: the Living Planet Index. *Conservation Biology* **23**:317–327.
- Cooper MW, Di Minin E, Hausmann A, Qin S, Schwartz AJ, Correia RA. 2019. Developing a global indicator for Aichi Target 1 by merging online data sources to measure biodiversity awareness and engagement. *Biological Conservation* **230**:29–36.
- Daniel TC, et al. 2012. Contributions of cultural services to the ecosystem services agenda. *Proceedings of the National Academy of Sciences of the United States of America* **109**:8812–8819.
- Fernández-Bellón D, Kane A. 2020. Natural history films raise species awareness—a big data approach. *Conservation Letters* **13**:e12678.
- Fink C, Hausmann A, Di Minin E. 2020. Online sentiment towards iconic species. *Biological Conservation* **241**:108289.
- Food and Agriculture Organization (FAO). 2020. ASFIS list of species for fishery statistics purposes. FAO, Rome.
- Fukano Y, Tanaka Y, Soga M. 2020. Zoos and animated animals increase public interest in and support for threatened animals. *Science of the Total Environment* **704**:135352.
- Hallmann CA, et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE* **12** (e0185809). <https://doi.org/10.1371/journal.pone.0185809>.
- IPBES. 2016. The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.
- IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services). 2019. The global assessment report on biodiversity and ecosystem services — summary for policymakers. IPBES, Bonn.
- IUCN (International Union for Conservation of Nature). 2020. IUCN Red List of threatened species. Version 2020–2. IUCN, Gland, Switzerland. Available from [www.iucnredlist.org](http://www.iucnredlist.org) (accessed March 2020).
- Jarić I, Bellard C, Courchamp F, Kalinkat G, Meinard Y, Roberts DL, Correia RA. 2020. Societal attention toward extinction threats: a comparison between climate change and biological invasions. *Scientific Reports* **10**:11085.
- Ladle RJ, Correia RA, Do Y, Joo GJ, Malhado AC, Proulx R, Roberge JM, Jepson P. 2016. Conservation culturomics. *Frontiers in Ecology and the Environment* **14**:269–275.
- Ladle RJ, Jepson P, Correia RA, Malhado AC. 2019. A culturomics approach to quantifying the salience of species on the global internet. *People and Nature* **1**:524–532.
- Leadley PW. 2013. Progress towards the Aichi Biodiversity Targets: an assessment of biodiversity trends, policy scenarios and key actions. *Global Biodiversity Outlook 4*. Convention on Biological Diversity, Montreal.
- Lenda M, Skórka P, Mazur B, Sutherland W, Tryjanowski P, Moroń D, Meijaard E, Possingham HP, Wilson KA. 2020. Effects of amusing memes on concern for unappealing species. *Conservation Biology* **34**:1200–1209.
- Loh J, Green RE, Ricketts T, Lamoreux J, Jenkins M, Kapos V, Randers J. 2005. The Living Planet Index: using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **360**:289–295.
- Mace GM, Barrett M, Burgess ND, Cornell SE, Freeman R, Grooten M, Purvis A. 2018. Aiming higher to bend the curve of biodiversity loss. *Nature Sustainability* **1**:448–451.
- Mccallum ML, Bury GW. 2013. Google search patterns suggest declining interest in the environment. *Biodiversity and Conservation* **22**:1355–1367.
- Mcowen CJ, Ivory S, Dixon MJR, Regan EC, Obrecht A, Tittensor DP, Teller A, Chenery AM. 2016. Sufficiency and suitability of global biodiversity indicators for monitoring progress to 2020 targets. *Conservation Letters* **9**:489–494.
- McRae L, Deinet S, Freeman R. 2017. The Diversity-weighted Living Planet Index: controlling for taxonomic bias in a global biodiversity indicator. *PLOS ONE* **12** (e0169156). <https://doi.org/10.1371/journal.pone.0169156>.

- Millard JW, Freeman R, Newbold T. 2020. Text-analysis reveals taxonomic and geographic disparities in animal pollination literature. *Ecography* **43**:44–59.
- Mittermeier JC, Roll U, Matthews TJ, Grenyer R. 2019. A season for all things: phenological imprints in Wikipedia usage and their relevance to conservation. *PLOS Biology* **17**:e3000146.
- Papworth SK, Nghiem TPL, Chimalakonda D, Posa MRC, Wijedasa LS, Bickford D, Carrasco LR. 2015. Quantifying the role of online news in linking conservation research to Facebook and Twitter. *Conservation Biology* **29**:825–833.
- Powney GD, Carvell C, Edwards M, Morris RK, Roy HE, Woodcock BA, Isaac NJ. 2019. Widespread losses of pollinating insects in Britain. *Nature Communications* **10**:1018.
- R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Roberge JM. 2014. Using data from online social networks in conservation science: which species engage people the most on Twitter? *Biodiversity and Conservation* **23**:715–726.
- Roll U, Mittermeier JC, Diaz GI, Novosolov M, Feldman A, Itescu Y, Meiri S, Grenyer R. 2016. Using Wikipedia page views to explore the cultural importance of global reptiles. *Biological Conservation* **204**:42–50.
- Rosindell J, Wong Y. 2020. OneZoom. Available from <https://www.onezoom.org/> (accessed March 2020).
- Scheffers BR, Oliveira BF, Lamb I, Edwards DP. 2019. Global wildlife trade across the tree of life. *Science* **366**:71–76.
- Sherren K, Smit M, Holmlund M, Parkins JR, Chen Y. 2017. Conservation culturomics should include images and a wider range of scholars. *Frontiers in Ecology and the Environment* **15**:289–290.
- Smith TJ, Saunders ME. 2016. Honey bees: the queens of mass media, despite minority rule among insect pollinators. *Insect Conservation and Diversity* **9**:384–390.
- Soriano-Redondo A, Bearhop S, Lock L, Votier SC, Hilton GM. 2017. Internet-based monitoring of public perception of conservation. *Biological Conservation* **206**:304–309.
- UNEP CBD (United Nations Environment Programme Convention on Biological Diversity. 2010. Strategic plan for biodiversity 2011–2020. CBD, Montreal.
- Union for Ethical BioTrade (UEBT). 2019. Biodiversity barometer. UEBT, Amsterdam.
- Verissimo D, Anderson S, Tlustý M. 2020. Did the movie Finding Dory increase demand for blue tang fish? *Ambio* **49**:903–911.
- Wong Y, Rosindell J. 2020. Dynamic visualisation of million-tip trees: the OneZoom project. Preprint. bioRxiv: 2020.10.14.323055.

