Article:

Johnson, T.F. orcid.org/0000-0002-6363-1825, Cornford, R., Dove, S. et al. (2 more authors) (2023) Achieving a real-time online monitoring system for conservation culturomics. Conservation Biology, 37 (4). e14096. ISSN 0888-8892

https://doi.org/10.1111/cobi.14096

Received: 24 October 2022 | Revised: 3 February 2023 | Accepted: 17 March 2023

DOI: 10.1111/cobi.14096

Conservation Biology

METHOD

Achieving a real-time online monitoring system for conservation culturomics

Thomas Frederick Johnson[1] | **Richard Cornford**[2,6] | **Shawn Dove**[2,3] | **Robin Freeman**[2] | **Joseph Millard**[4,5]

[1]School of Biosciences, University of Sheffield, Sheffield, UK

[2]Institute of Zoology, Zoological Society of London, London, UK

[3]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK

[4]Department of Life Sciences, Natural History Museum, London, UK

[5]Leverhulme Centre for Demographic Science, Nuffield College, University of Oxford, Oxford, UK

[6]Biodiversity, Ecology and Conservation Group, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Correspondence
Joseph Millard, The Natural History Museum, London, UK.
Email: joseph.millard@nhm.ac.uk

Article impact statement: A real-time Species Awareness Index for culturomics can inform conservation policy through monitoring of human interactions with nature.

Abstract
Environmental monitoring is increasingly shifting toward a set of systems that describe changes in real time. In ecology specifically, a series of challenges have prevented the roll-out of real-time monitoring for features such as biodiversity change or ecosystem service provision. Conservation culturomics, a field concerned with interactions between people and nature, is well placed to demonstrate how monitoring might move toward a network of real-time platforms, given its existence exclusively in the digital realm. We examined a set of considerations associated with the development of real-time monitoring platforms for conservation culturomics and introduce a near real-time platform for the Species Awareness Index, a global index of changing biodiversity awareness derived from the rate of change in page views for species on Wikipedia. This platform will update automatically each month, operating in near real time (https://joemillard.shinyapps.io/Real_time_SAI/). There are plans to make the underlying data queryable via an application programing interface independent of the platform. The real-time Species Awareness Index will represent the first real-time and entirely automated conservation culturomic platform and one of the first real-time platforms in the discipline of ecology. Real-time monitoring for culturomics can provide insight into human–nature interactions as they play out in the physical realm and provide a framework for the development of real-time monitoring in ecology. Real-time monitoring metrics can be processed on private virtual machines and hosted on publicly available cloud services. Conservation now needs an online, real-time observatory that can evolve with the structure of the web.

KEYWORDS
automated monitoring, conservation culturomics, real-time monitoring, Species Awareness Index, Wikipedia

Obtención de un sistema virtual de monitoreo en tiempo real para la culturomía de la conservación

Resumen: El monitoreo ambiental se enfoca cada vez más en un conjunto de sistemas que describen los cambios en tiempo real. En cuanto a la ecología, una serie de obstáculos ha impedido el despliegue del monitoreo en tiempo real para funciones como el cambio en la biodiversidad o el suministro de servicios ambientales. La culturomía de la conservación, un campo enfocado en las interacciones entre las personas y la naturaleza, es una buena opción para demostrar cómo el monitoreo podría transformarse en una red de plataformas en tiempo real, dado que sólo existe en el ámbito digital. Analizamos una serie de consideraciones asociadas con el desarrollo de plataformas de monitoreo en tiempo real para la culturomía de la conservación e introdujimos una plataforma casi en tiempo real para el Índice de Conciencia de Especies, un índice mundial del cambio en la conciencia sobre la biodiversidad derivado de la tasa de cambio en las visitas a las páginas de Wikipedia de diferentes especies. Esta plataforma se actualizará automáticamente cada mes, por lo que opera casi en tiempo real (https://joemillard.shinyapps.io/Real_time_SAI/). Existen planes para hacer que los datos subyacentes sean consultables por medio de una interfaz de programación de aplicaciones independiente de la plataforma. El Índice de Conciencia de Especies en tiempo real será la primera plataforma de culturomía de la conservación automatizada por completo y en tiempo real, así como una de las primeras plataformas de este tipo para la disciplina de la ecología. El monitoreo en tiempo real de la culturomía puede proporcionar información sobre las interacciones humano-naturaleza conforme se desarrollan en el ámbito físico y también un marco de trabajo para el desarrollo del monitoreo en tiempo real para la ecología. Las medidas del monitoreo en tiempo real pueden procesarse en máquinas virtuales privadas y albergarse en servicios públicos de nubes de almacenamiento. Hoy en día, la conservación necesita un observatorio en línea y en tiempo real que pueda evolucionar con la estructura de la web.

PALABRAS CLAVE

culturomía de la conservación, índice de conciencia de especies, monitoreo automatizado, monitoreo en tiempo real, Wikipedia

【摘要】

环境监测正在日益向能够实时描述变化的系统转变。特别是在生态学方面,一系列挑战阻碍了开发生物多样性变化或生态系统服务供给等特征的实时监测工具。保护文化组学是一个关注人与自然相互作用的领域,它完全存在于数字世界,因此可以很好地展示监测如何向实时平台网络的方向发展。本研究分析了与保护文化组学实时监测平台发展相关的一系列考虑因素,并为物种关注度指数 (Species Awareness Index) 设计了一个接近实时监测的平台。物种关注度指数是一个关于生物多样性意识变化的全球指数,来自维基百科上物种页面浏览量的变化情况。这个平台将每月自动更新,以接近实时的方式运作 (https://joemillard.shinyapps.io/Real_time_SAI/) 。我们计划通过一个独立于平台的应用程序接口来提供基础数据查询。实时物种关注度指数将代表第一个实时和完全自动化的保护文化组学平台,这也是生态学中第一批实时平台之一。文化组学的实时监测可以为人类与自然的相互作用提供深入见解,因为它们存在于现实世界,并为生态学实时监测的发展提供了框架。实时监测指标可以在私人虚拟机上处理,并在公共云服务平台上托管。我们认为,保护工作现在需要一个在线的、实时的、随网络结构发展而演化的观测站。**【翻译: 胡怡思; 审校: 聂永刚】**

关键词: 保护文化组学, 实时监测, 物种关注度指数, 自动化监测, 维基百科

INTRODUCTION

Real-time monitoring has revolutionized environmental management, offering insight and foresight on the risk of natural and anthropogenic disasters (Smith et al., 2017) and influencing human health and disease spread (Hadfield et al., 2018). Real-time ecological and biodiversity monitoring could potentially offer similar benefits, but has historically been constrained by challenges in wide-scale manual data collection (Biber, 2013), as well as a lack of infrastructure and expertise for automating analyses and reporting. However, the development of monitoring approaches, such as eDNA (Garlapati et al., 2019), remote sensing (Steenweg et al., 2017), acoustics (Sethi et al., 2020), animal satellite telemetry (Jetz et al., 2022; Wall et al., 2014), and culturomics (Ladle et al., 2016), offers potential pathways toward real-time monitoring of biodiversity and ecosystem services.

Conservation culturomics, a subfield of culturomics dedicated to the study of human relationships with nature and wildlife (Ladle et al., 2016), is ideally placed for the development of real-time monitoring systems, given the required data are found in the digital realm. The study of conservation culturomics has improved understanding of human–nature interactions. For instance, conservation culturomics has shown that interest in biodiversity increased during COVID-19 lockdowns (Roll et al., 2021) and that interest in nature changes according to season (Mittermeier et al., 2019). From a conservation perspective, culturomics has revealed patterns of wildlife trade (Li & Hu, 2021) and helped to gather information on

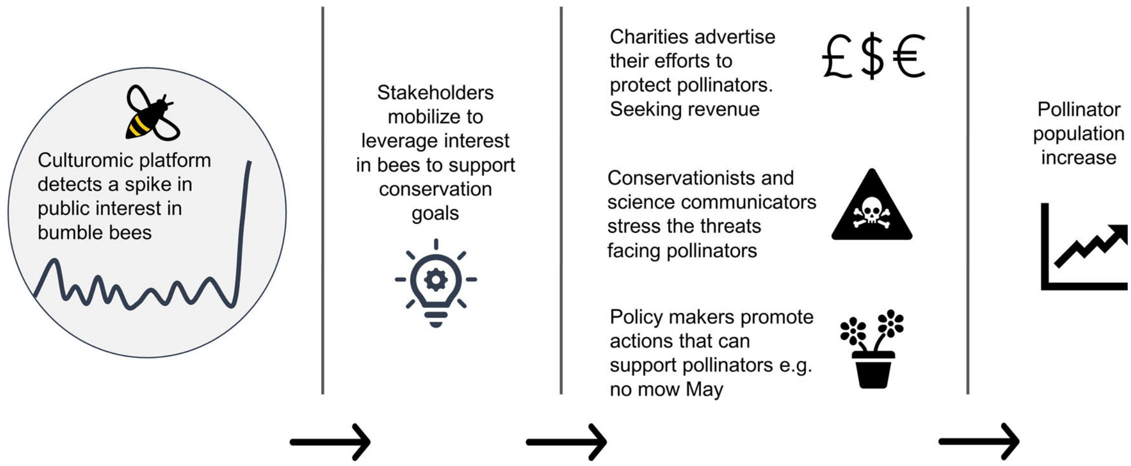

FIGURE 1 Potential application and impact of real-time culturomics, depicting how it could be possible to leverage spikes in interest to support conservation action.

wildlife-associated recreational activities (Monkman et al., 2018; Otsuka & Yamakoshi, 2020). All of the above are changing the conservationist's understanding of human–nature interactions.

Online conservation culturomics platforms have previously been developed for awareness of biodiversity (Caetano et al., 2021; Cooper et al., 2019), but these lack updates in real time, meaning they represent a static snapshot of public awareness. This introduces a lag between data collection and resulting publication, sometimes taking years; it is possible that results become outdated by the time the publication is released. However, even if the results still hold after years in the publication process, arguably, conservation culturomics insight is of greatest value in the immediate term (i.e., in real time), so key players driving conservation action can react to current opinion.

Real-time monitoring of awareness or interest in species could offer actionable insight. For example, if interest in a species spikes, policy makers and conservationists could target information on how people can help support this species. Further, charities could target advertising toward this species to gain increased support and funding, building on the traction of temporarily high-profile species (Figure 1). Outside conventional conservation culturomics, spikes in culturomic activity (e.g., Twitter posts, Wikipedia searches, citizen science records) could reveal protected areas at risk (Guedes-Santos et al., 2021) or ecological phenomena and help address conservation issues. For instance, a mobile app called Hornet Watch is currently used for reporting sightings of the invasive Asian Hornet in the United Kingdom, but this could be further enhanced by incorporating passive culturomic data from social media, enabling the rapid implementation of management to prevent further spread of an invasive species (CEH, 2017). Incorporating passive data (e.g., insight into human–nature interaction extracted from general sources like social media) into alert systems like Hornet Watch could help target limited conservation resources, making real-time culturomics platforms of great potential value to the field of conservation.

We devised a set of considerations in building a real-time conservation culturomics platform and introduce a near real-time platform for the Species Awareness Index as a case study. The Species Awareness Index describes the rate of change in Wikipedia page views for approximately 40,000 species across 10 Wikipedia languages, adjusted for the background change in each language (Millard et al., 2021). This platform is hosted online as a Shiny app (https://joemillard.shinyapps.io/Real_time_SAI/), which updates automatically each month through a batch process that runs on a virtual machine. This real-time platform is cost-effective, scalable, and implementable with modest programing skills. We considered a future for real-time monitoring in the context of conservation culturomics and the need for long-term collaborative thinking.

CONSIDERATIONS FOR REAL-TIME CONSERVATION CULTUROMICS

There are many challenges associated with the development of real-time monitoring platforms for conservation culturomics: platforms need to track robust metrics that have a meaningful interpretation in real-time; platform hosting needs to be cost-effective and computationally efficient; platforms need to be built such that they account for the structure of application programing interfaces (APIs) in terms of limits, permissions, and potential changes; and platforms need to have a long-term funding and maintenance plan, ideally linked to a particular conservation or natural history institution with an in-house informatics team. Many of the conservation culturomics considerations we examined will also be valuable for any real-time monitoring program in ecology.

Robust metrics of human–nature interactions

A prerequisite for any real-time conservation culturomics platform is the development of robust metrics that take some web-derived data sources as an input and then outputs a metric with a temporal dimension that meaningfully describes some form of human–nature interaction. Correia et al. (2021) summarize the general requirements for a conservation culturomics metric, so we focused on the properties required for real-time utility. Conservation culturomic metrics can be highly abstracted describing human–nature interactions in a coarse manner or specifically oriented toward a particular population or behavior. Highly abstracted metrics include those that aim to measure biodiversity awareness (Caetano et al., 2021; Cooper et al., 2019; Millard et al., 2021) or broad engagement with nature (Phillips et al., 2022; Roll et al., 2021). Specifically oriented metrics include those that track interactions between humans and particular species at a high taxonomic or geographic resolution (Acerbi et al., 2020; Sbragaglia et al., 2021).

For either highly abstracted or highly resolved human–nature metrics, the temporal dimension is particularly important in real-time monitoring for 2 key reasons. First, the metric needs to be measured at a resolution sufficient to observe heterogeneity in the phenomena of interest. For example, seasonality of human interest in biodiversity will not be detected if the culturomics data of interest are measured at an annual resolution. Some sources, such as Twitter data, are time-stamped at a very high resolution (seconds) (Twitter, 2022a), making them highly amenable to understanding high-resolution real-time insights, whereas other data sources are aggregated at the daily resolution (Wikimedia, 2022). Second, real-time metrics must be feasibly derivable at the temporal resolution of interest. Each metric must be at least as fast to derive as its data are collected (i.e., if data are collected hourly, but it takes more than an hour to derive metrics for a given selection of data, the real-time metric will begin to lag and become outdated). This may be problematic when working with high-dimensional visual, text, or audio data in which a series of computationally intensive steps may be required to process the data before the derivation of any metric.

Unlike the publication of a conservation culturomics metric in the academic literature, a real-time platform is uniquely placed such that it can be updated according to current thinking in the academic community. This is particularly relevant to metrics such as the Species Awareness Index (Millard et al., 2021), which leans heavily on the methodological backbone of the Living Planet Index. Although the Living Planet Index has undergone multiple improvements (Collen et al., 2009; McRae et al., 2017), it is still subject to criticism and suggestions for further improvement (Leung et al., 2020; Puurtinen et al., 2022). Moreover, metrics such as the Species Awareness Index (Millard et al., 2021) or Biodiversity Engagement Indicator (Cooper et al., 2019) have not yet been sufficiently critiqued for full understanding of the extent to which they are useful or meaningful. As in the Living Planet Index (Ledger et al., 2022), the authors of the Species Awareness Index are receptive to feedback to ensure that the field continues to evolve. In the manner of living reviews (Elliott et al., 2014), updatable platforms that can provide a robust current account of human–nature interactions should be encouraged. For a full description of Species Awareness Index limitations, see Millard et al. (2021).

Hosting a real-time platform

A real-time platform could be hosted on any website, but it often requires specialized html programing skills, which are uncommon among conservation culturomics researchers and practitioners (Hampton et al., 2017). This skill gap could be resolved by employing web developers, but this would inflate the cost of platform development and maintenance. A practical solution is to use established dashboard platforms like R Shiny apps (Chang et al., 2023). Shiny apps are developed using R, and R is a common programing language among quantitative conservationists (Lai et al., 2019); thus, apps can be developed and maintained by conservation culturomics researchers. Shiny apps are also cost-effective because the price depends on their use. For high use, Shiny applications may struggle with high traffic volume, in which case it could be worthwhile exploring services such as ShinyProxy, which are better equipped for concurrent application use (Open Analytics, 2022).

Although Shiny applications are suitable for hosting a real-time platform, they are ill-equipped for the intensive computation needed to derive conservation culturomic metrics. A solution to this could be to use virtual machines or servers as a back end to derive the metric and store associated data and then to use Shiny as the front end for hosting the metric. Examples of servers include Amazon web services (AWS, 2022), Microsoft Azure (Microsoft, 2022), and Google Cloud (Google Cloud, 2022). These servers are already widely used in high-performance computing applications (e.g., deep-learning [Jauro et al., 2020]) and are cost-effective because charges are generally proportional to use. By shifting all computation to online servers, instead of local computers, there is less risk of disruption from software updates or power outages. Data storage on virtual machines could mean writing .rds or .csv files to disk on the machine's hard drive or writing monthly updates to a structured relational database (e.g., MySQL or PostgreSQL). Writing monthly updates to disk as separate .csv files means a lower overhead on development and technical skills, but it is more susceptible to data loss or overwriting and less efficient for storage. Perhaps the biggest obstacle to using online servers is researcher's unfamiliarity with them; they can be dense and complex to use. But, making this shift would increase the resilience of platforms and improve their longevity.

Accounting for the structure of APIs

The data used in conservation culturomics metrics are often drawn from APIs (defined above), which allow online-hosted data sets to be queried and downloaded. These APIs are dynamic in nature because they host a constant stream of new data. They are also subject to changing terms of use, can alter the format data are provided in, and can shift download limits (often called rate limits, meaning the rate at which data can be requested from the API). These changes in data usage rights and format could cause a platform to fail. It is likely a program of long-term maintenance would be needed to address these problems. It is also important that the community of researchers using and maintaining these real-time platforms develop relationships with and receive communication from data providers, to foresee and address changes before the platform fails.

Although API changes will need to be addressed to prevent platforms from failing, it is worth noting that changes in API use will not always be restrictive. For example, Twitter has recently changed their policy to boost API access for academics (Twitter, 2022a). Previously researchers could only download a small selection of recent Tweets, but these restrictive download limits were then removed for academic research purposes, enabling access to the full catalogue, before being quickly reinstated in February 2023. This volatility in access presents challenges and opportunities for conservation culturomics dashboards to not just monitor real-time changes, but also historical change.

Caution needs to be applied when using online culturomics data sets that store human–nature data. From a practical and legal perspective, websites such as Wikipedia and Twitter have strict API restrictions, in terms of the purpose, regularity, pattern, and quantity of requests. If a user (IP address) breaks these terms, they risk being blocked, potentially permanently (Twitter, 2022b). Further, the licenses for using data from websites like Wikipedia and Twitter often prevent the sharing of raw data and only allow the presentation of aggregated data. Finally, although the data are publicly available, there are still ethical considerations for using these data sets (Di Minin et al., 2021). Websites with personal profiles and data such as Twitter or Flickr should be used carefully. It is important that any culturomics application, real time or not, abide by the highest ethical and legal standards (Correia et al., 2021; Thompson et al., 2021).

Developing a funding plan for the long-term

We propose that a clear, consistent, and long-term funding plan (i.e., depicting the resources needed to ensure sufficient personnel for the long-term improvement and maintenance of the platform) be developed to allow real-time conservation culturomics platforms to run with longevity. It is also important that funding be available to handle the cost associated with virtual machines, servers, and Shiny apps, with contingency planning in place to handle changing prices. Given these funding requirements and the uncertainty associated with academic funding streams, academia may not be the best place for real-time platforms to be maintained. One option could be to host the real-time platform with a conservation or natural history organization that has an established informatics team and sits within a network of practitioners that could make use of the platform (e.g., Royal Society for the Protection of Birds [RSPB], Zoological Society of London [ZSL], Birdlife International, International Union for the Conservation of Nature [IUCN], UN Environment Programme World Conservation Monitoring Centre [UNEP-WCMC], natural history museums). Perhaps the ideal location would be to embed conservation culturomics platforms within the framework of essential biodiversity variables (e.g., moving from solely tracking changes in biodiversity to also tracking changes in people's relationship with biodiversity; currently led by the Group on Earth Observations Biodiversity Observation Network [GEO-BON]).

A NEAR REAL-TIME SPECIES AWARENESS INDEX

To provide a practical example of a real-time conservation culturomics metric, we converted the static Species Awareness Index into a (near) real-time product. The Species Awareness Index describes the dynamics of Wikipedia page views for more than 40,000 species, offering a unique insight into public interest across taxonomic groups over time. The Species Awareness Index is designed for use at global scales, lacks spatial information, and has a wide array of potential uses and users. With up-to-date information, the Species Awareness Index could be incorporated into global monitoring initiatives, such as GEO-BON's essential biodiversity variables, or used by governmental and nongovernmental organizations to track changing interest in biodiversity at a global scale. This is a critical gap in existing extinction risk indices like the Living Planet Index because fluctuating interest in biodiversity will likely alter realized biodiversity patterns (e.g., if interest in biodiversity increases, this could have a downstream effect on biodiversity change). The Species Awareness Index could also be used to track the societal extinction of species (i.e., species vanishing from our collective memories, knowledge, and cultures) (Jarić et al., 2022), as a complement to biological extinction measured through metrics such as the IUCN Red List Index (Butchart et al., 2007), the Living Planet Index (Collen et al., 2009), or the Biodiversity Intactness Index (De Palma et al., 2021) (Figure 2).

With the Species Awareness Index, there is also the potential to conduct pre- and poststudies on interventions (e.g., release of the IPBES report [IPBES, 2019]) and events (e.g., media attention following the hunting of Cecil the Lion) and tailor communication strategies accordingly. For instance, if global interest in large carnivores spikes, this could be the prime opportunity to educate the public on why these species are threatened and how the behavior of people can help protect them. Converting the static Species Awareness Index into a near real-time version also addresses an existing Species Awareness Index limitation, shared by other conservation culturomics indices. Specifically, the static Species Awareness Index has not been updated since early 2020, in a period when dynamics may be highly volatile (e.g., changing relationships with nature during the pandemic [e.g., Roll et al., 2021]). Other limitations exist (Millard et al., 2021).

For other conservation culturomics metrics with the added spatial component, the use cases are even greater. For instance, a spike in interest of 1 taxonomic group and 1 location (e.g., return of migrant birds to Europe) could be an opportune moment to seek conservation funds to protect these threatened species. One could argue that many ecological phenomena are cyclical and seasonal, simply following the phenology of species ecology, and so a real-time system is of little value. However,

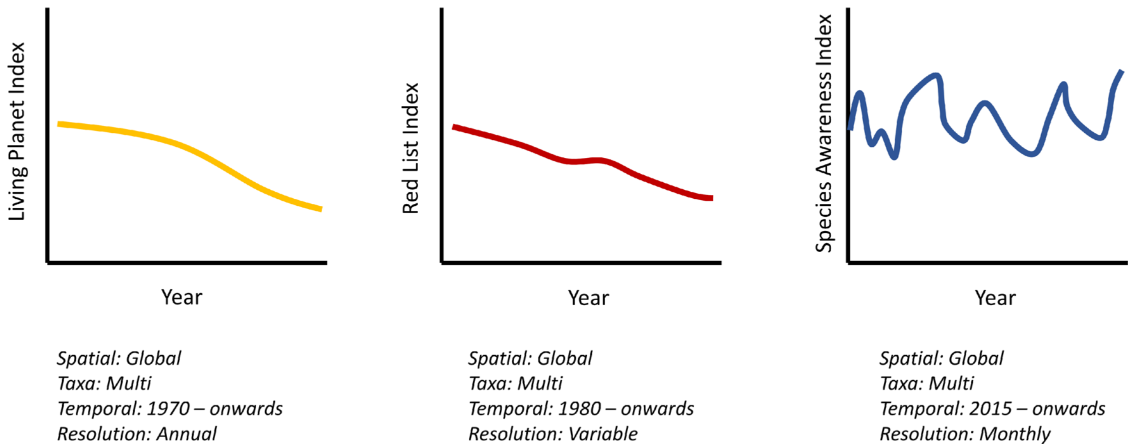

FIGURE 2 The Species Awareness Index as an essential biodiversity variable. The index (right panel) could accompany other near-real indices, such as the International Union for the Conservation of Nature Red List Index (middle panel [Butchart et al., 2007]) and Living Planet index (left panel [Collen et al., 2009]), to describe changing extinction risk and wildlife population trends, respectively.

this is a flawed argument because people will not have the same interest in all phenological phenomena, so tracking real-time conservation culturomics can help target particular areas that are high profile; phenology is variable among taxa, time, and space (e.g., migrant birds can return anytime from February to May); and real-time conservation culturomics presents the possibility of tracking irregular and extreme events (e.g., Cecil the Lion).

The near real-time Species Awareness Index is a prototype. Development will inevitably be a continual process, as bugs are overcome and new features added. Its development was a challenging endeavor. Much of the discussion on development assumes the reader is familiar with the Species Awareness Index itself (Millard et al., 2021).

Changes to the core Species Awareness Index code

In the original Species Awareness Index publication, much of the code we wrote relied on the R package rLPI (Institute of Zoology, Zoological Society of London, 2022). The rLPI is a package designed specifically for the calculation of the Living Planet Index (LPI), an aggregated trend representing the average change in abundance of many vertebrate populations. Given that the Species Awareness Index is inspired by the LPI, it made sense at the initial stage to apply it to Wikipedia views. For a couple of reasons, we moved away from the rLPI package and instead developed a set of scripts specifically for the Species Awareness Index. First, to be amenable for use with the rLPI package, Wikipedia view data needed to be reformatted to look like a set of population trends. With reformatting, the more informative Wikipedia-specific column names can be used. Second, rLPI contains a large quantity of additional code that we did not require in the Species Awareness Index. For example, rLPI automatically bootstraps species page trends before they are adjusted for change in a random set of pages.

In addition to the changes above, at present we do not jackknife the bootstrapped trend of random-adjusted species pages or include a weighting. In the original version of the code, bootstrapped trends were jackknifed by language, with influential languages then removed from the overall trend. Previously, this resulted in the removal of the French language from the overall trend and the trend broken down by taxonomic class. Our eventual intention is to automatically jackknife the trend by language each month, calculate some parameter that summarizes the influence of individual languages, and then remove any language that surpasses a particular threshold of influence. For now, however, this has not been implemented, but represents a priority for further development. Currently, users should be cautious in interpreting the overall and class-level indices because these are likely heavily influenced by change in the French language (see Millard et al. [2021] for an explanation as to why this is the case). In the Supporting Information of our original paper (Millard et al., 2021), we explored a set of weightings by total internet users and the number of species in which a language is represented. Again, for both these weightings, we do not yet have an implementation in our real-time version, but we intend to implement both.

Building the platform interface

The real-time Species Awareness Index is hosted online on a Shiny app (https://joemillard.shinyapps.io/Real_time_SAI/), which updates automatically through monthly writes to an AWS droplet. We used this approach because Shiny is written explicitly to integrate with R. Given that the Species Awareness Index code is written in R, it therefore made sense to build the platform in Shiny because we could just carry over all the ggplot2 code (Wickham et al., 2022) used to build the visualizations on the trends page. At the moment, the platform is hosted at shinyapps.io under RStudio's free hosting. Although free hosting is sufficient for a prototype, our long-term intention is to

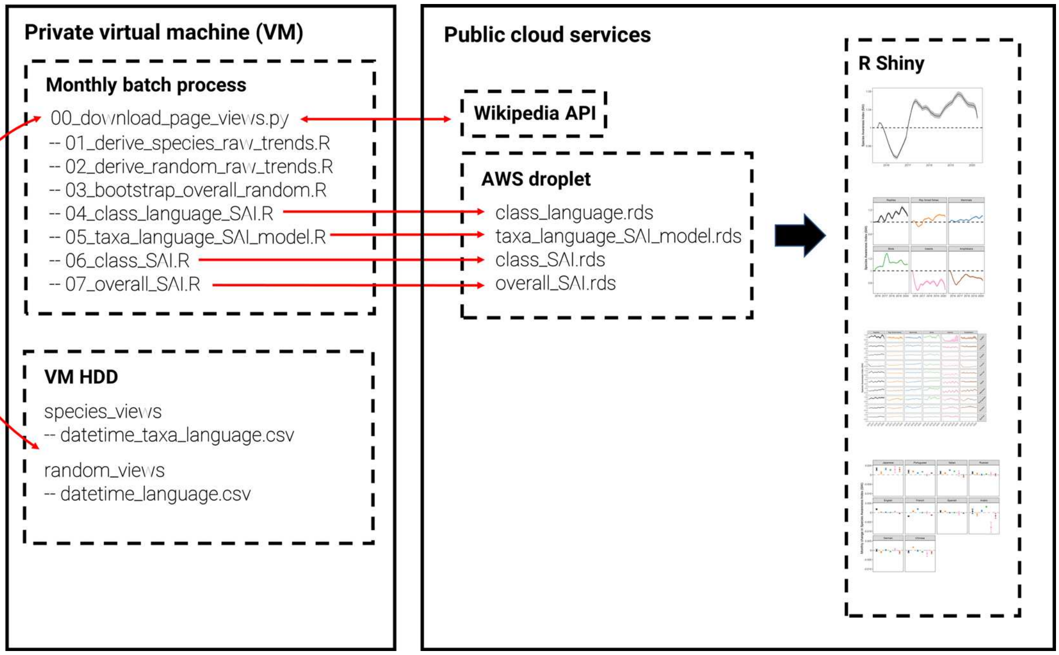

FIGURE 3 Schematic summary of the real-time Species Awareness Index pipeline that runs on a virtual machine each month (right, software running on a virtual machine; left, publicly available cloud service software). Each month a new set of views from the Wikipedia API (according to the date of the previous set) downloads to the virtual machine (VM) and the daily average views per month for the random and species pages are calculated. These files are written to the virtual machine hard disk drive (HDD) (script 00). That initial Python script runs 7 other scripts that result in 4 rds files, each of which is written to an Amazon Web Services (AWS) droplet on the cloud and then read into R Shiny whenever the platform is visited. The Shiny platform can then depict the key outputs from the Species Awareness Index (4 graphs in right-most panel, the contents of which are not to be interpreted within this figure).

shift to ShinyProxy, which enables enterprise-level traffic handling under an open-source model (Open Analytics, 2022). Such a change will require bundling the platform into an R package, which can then be installed via a Docker image (Docker, 2022).

Summary of the whole real-time Species Awareness Index pipeline

Although R Shiny is convenient as a tool due to its integration with R, this close integration makes it tempting to run large quantities of R code as each instance of the app is initiated. In the case of the Species Awareness Index, this would mean running a pipeline that takes a number of hours each time a user wants to view a specific trend. Such a time lag before viewing the Species Awareness Index trends would render the platform unusable. Therefore, we opted to shift all core Species Awareness Index code and Wikipedia view downloads off the app, with this running in the background on a virtual machine in a batch Python process each month. The output of this batch process is then written to an AWS droplet from which the Shiny app reads each time the URL is visited by a user. By building the platform in this way, Shiny will always request from the most recent version of the Species Awareness Index output, and it will only ever take as long to load as it does to generate the set of ggplots on the trends page. Below we describe each core step of the real-time Species Awareness Index pipeline that runs each month (Figure 3 contains a schematic of each core step in the pipeline). We do not describe the structure of our Shiny app itself because this software functions solely to visualize the data outputted by the monthly batch process.

In the first step, each month a Python script downloads Wikipedia views from each of the species and random pages previously included in the original Species Awareness Index publication (see Millard et al., 2021) and calculates an average number of daily views per month for each page. The underlying taxonomy and main Wikipedia page name for each of these species was taken from a Onezoom download (Wong & Rosindell, 2022), and the random pages were identified using the Random Page Wikipedia API (as in Millard et al. [2021]). We did not download a different set of random pages each month for each language because this would represent significant additional overhead, even without considering the need to download new time series for the whole period with each passing month. In the long term, our intention is to further test the influence of random page selection and potentially include a download of new complete series random pages each month.

To ensure that at each month the new download starts from the previous month, at the top of our batch Python script we read in the most recent version of the overall Species Awareness Index trend (from the local directory of a virtual machine). This file contains a date column containing all dates from the beginning to the current end of the time series. The Wikipedia API download then starts from the last month of that date column to the most recent month beyond that date column. Following the download completion of each taxonomic class in each language

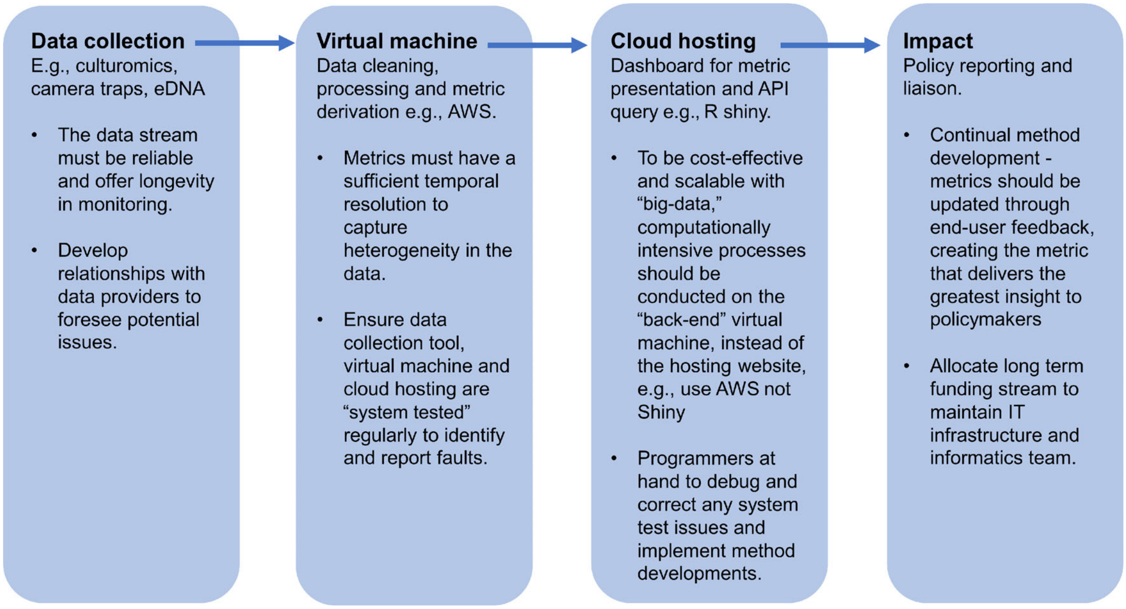

FIGURE 4 Four stages and considerations in each stage in a general framework for development of a real-time monitoring program in ecology (AWS, Amazon Web Services; IT, information technology).

and each random page in each language, all of these average daily views per month are then written to disk on a virtual machine with a unique time-stamped file name.

After the Wikipedia view download is complete, that same Python script then executes a set of R scripts on the virtual machine that run all of the relevant Species Awareness Index code. This batch process consists of 7 scripts that run in sequence: first, derivation of individual species page view trends; second, derivation of individual random page view trends; and third, bootstrapping of the random page view trend in each language. The fourth, fifth, sixth, and seventh scripts adjust change in species page views with change in a random set of pages and then generate each of the data files underpinning each of the graphs on the trends page of the platform. These 4 files are stored in an AWS droplet from which the Shiny app reads each time the platform is engaged. Each of these files are also made available for download on the download page of the platform. Our reasoning for hosting only these Shiny inputs on AWS is in part to keep cloud storage costs down, but also to ensure that the Shiny app can access these files once it has been deployed (i.e., code running on our virtual machine can access local directories on that virtual machine, but a deployed Shiny app cannot). In the future, our intention is to pivot our data storage to some form of SQL database with an associated API, which the Shiny app would request from at start-up. Researchers would then also be able to request subsets of random-adjusted species trends, for particular species, groups of species, languages, or time periods.

ENVISIONING A FUTURE FOR REAL-TIME MONITORING IN ECOLOGY

In the future, we hope that real-time monitoring platforms of the sort we built can sit within an online observatory of similar metrics. Such a platform would help to realize the aim of many in the conservation culturomics community (Jarić et al., 2020): an observatory that tracks analogues of societal extinction (Jarić et al., 2022). Ideally, such a monitoring system would incorporate a number of data sets, including Wikipedia, Baidu-Baike, Twitter, and Google, to capture a broad demographic and geographic distribution of users. An important consideration, however, is ensuring derived metrics are useful. Given real-time metrics carry a cost of development and maintenance, it is vital that developed metrics add value. The conservation culturomics researcher community should work closely with external stakeholders to codesign the next generation of conservation culturomics indices. Regardless of these problems, however, a real-time digital observatory is feasible because the underlying infrastructure in leveraging a real-time platform will be the same irrespective of data source.

Developing a real-time observatory of conservation culturomics data is a natural first step because an extensive array of data sources exist in the digital realm and so are amenable to the automated workflow we presented. However, this workflow could also form the basis of a general framework for real-time monitoring in ecology, defined by 4 stages: data are collected in real time; on a virtual machine, collected data are processed into metrics of change; metrics of change are visualized with a front-end tool such as R Shiny; and the front-end platforms deliver insight to policy makers in near real time (Figure 4). With the increasing development of automated data collection approaches in ecology, such as networked camera traps, which provide a continual stream of temporally and spatially resolved images (Wearn et al., 2017), there is a substantial opportunity for real-time monitoring in conservation culturomics and ecology more generally (Besson et al., 2022). Such a development would help realize a core aim of the CBD (Convention on Biology Diversity): a set of ambitious and modern indicators that are compiled and updated regularly (UNEP CBD, 2021).

SUMMARY

Given the large amounts of relevant digital data now available, conservation culturomics is ideally placed to lead the way in real-time monitoring. Our new platform is an ongoing project in which new features will be added and bugs fixed on a continual basis. Such a platform demonstrates the potential of conservation culturomics to provide insights on human–nature interactions as they play out in the physical realm. Our hope is that conservation culturomics researchers can come together to build a suite of real-time monitoring platforms that incorporate data from multiple online sources to help realize a core aim of the conservation culturomics community.

ACKNOWLEDGMENTS

First and foremost, we thank Wikipedia for providing freely accessible page view data via their APIs. We also thank Onezoom (J. Rosindell and Y. Wong) for supplying taxonomic reference data and the RSPB, NERC (GLiTRS, NE/V006800/1), the London NERC DTP, A. Purvis, R. Gregory, and M. Mills, who provided support, funding, or both during the development of this platform. Finally, we thank A. Cardwell for infographic development.

ORCID

Thomas Frederick Johnson https://orcid.org/0000-0002-6363-1825

REFERENCES

Acerbi, A., Kerhoas, D., Webber, A. D., McCabe, G., Mittermeier, R. A., & 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.

Amazon Web Services (AWS). (2022). Cloud computing services. https://aws.amazon.com/

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, *25*(12), 2753–2775.

Biber, E. (2013). The challenge of collecting and using environmental monitoring data. *Ecology and Society*, *18*(4), 68.

Butchart, S. H. M., Akçakaya, H. R., Chanson, J., Baillie, J. E. M., Collen, B., Quader, S., Turner, W. R., Amin, R., Stuart, S. N., & Hilton-Taylor, C. (2007). Improvements to the Red List Index. *PLoS ONE*, *2*, e140. https://doi.org/10.1371/journal.pone.0000140

Caetano, G., Roll, U., & Verissimo, D. (2021). *Measuring global awareness of nature*. On the Edge Conservation. https://doi.org/10.13140/RG.2.2.26970.06088

Centre for Ecology and Hydrology (CEH). (2017). *Asian hornet watch*. Author.

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). *shiny: Web application framework for R*. R package.

Collen, B., Loh, J., Whitmee, S., McRAE, L., Amin, R., & Baillie, J. E. M. (2009). Monitoring change in vertebrate abundance: The living planet index. *Conservation Biology*, *23*, 317–327.

Correia, R. A., Ladle, R., Jaric, I., Malhado, A., Mittermeier, J., Roll, U., Soriano-Redondo, A., Verissimo, D., Fink, C., Hausmann, A., Guedes-Santos, J., Vardi, R., & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, *35*(2), 398–411.

Cooper, M. W., Di Minin, E., Hausmann, A., Qin, S., Schwartz, A. J., & Correia, R. A. (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.

De Palma, A., Hoskins, A., Gonzalez, R. E., Börger, L., Newbold, T., Sanchez-Ortiz, K., Ferrier, S., & Purvis, A. (2021). Annual changes in the Biodiversity Intactness Index in tropical and subtropical forest biomes, 2001–2012. *Scientific Reports*, *11*, 20249.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*, 437–446.

Docker. (2022). Home - Docker. https://www.docker.com/

Elliott, J. H., Turner, T., Clavisi, O., Thomas, J., Higgins, J. P. T., Mavergames, C., & Gruen, R. L. (2014). Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine*, *11*, e1001603. https://doi.org/10.1371/journal.pmed.1001603

Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P., & Ramana Murthy, M. V. (2019). A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Reviews in Environmental Science and Biotechnology*, *18*, 389–411.

Google Cloud. (2022). Cloud Computing Services. https://cloud.google.com/

Guedes-Santos, J., Correia, R., Malhado, A., & Ladle, R. J. (2021). A digital approach to quantifying political vulnerability of protected areas. *Environmental Science & Policy*, *124*, 616–626.

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*, *34*(23), 4121–4123.

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R. R., Boettiger, C., Collins, S. L., Gross, L. J., Fernández, D. S., Budden, A., White, E. P., Teal, T. K., Labou, S. G., & Aukema, J. E. (2017). Skills and knowledge for data-intensive environmental research. *Bioscience*, *67*, 546–557.

Institute of Zoology, Zoological Society of London. (2022). *rlpi package (beta)*. Author.

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES Secretariat. https://doi.org/10.5281/zenodo.3831673

Jarić, I., Roll, U., Arlinghaus, R., Belmaker, J., Chen, Y., China, V., Douda, K., Essl, F., Jähnig, S. C., Jeschke, J. M., Kalinkat, G., Kalous, L., Ladle, R., Lennox, R. J., Rosa, R., Sbragaglia, V., Sherren, K., Šmejkal, M., Soriano-Redondo, A., … Correia, R. A. (2020). Expanding conservation culturomics and iEcology from terrestrial to aquatic realms. *PLoS Biology*, *18*, e3000935. https://doi.org/10.1371/journal.pbio.3000935

Jarić, I., Roll, U., Bonaiuto, M., Brook, B. W., Courchamp, F., Firth, J. A., Gaston, K. J., Heger, T., Jeschke, J. M., Ladle, R. J., Meinard, Y., Roberts, D. L., Sherren, K., Soga, M., Soriano-Redondo, A., Veríssimo, D., & Correia, R. A. (2022). Societal extinction of species. *Trends in Ecology & Evolution*, *37*, 411–419.

Jauro, F., Chiroma, H., Gital, A. Y., Almutairi, M., Abdulhamid, S. M., & Abawajy, J. H. (2020). Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend. *Applied Soft Computing*, *96*, 106582.

Jetz, W., Tertitski, G., Kays, R., Mueller, U., & Wikelski, M. (2022). Biological Earth observation with animal sensors. *Trends in Ecology and Evolution*, *37*(4), 293–298.

Ladle, R. J., Correia, R. A., Do, Y., Joo, G. J., Malhado, A. C., Proulx, R., Roberge, J. M., & Jepson, P. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment*, *14*(5), 269–275.

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, *10*, e02567. https://doi.org/10.1002/ecs2.2567

Ledger, S. E. H., McRae, L., Loh, J., Almond, R., Böhm, M., Clements, C. F., Currie, J., Deinet, S., Galewski, T., Grooten, M., Jenkins, M., Marconi, V., Painter, B., Scott-Gatty, K., Young, L., Hoffmann, M., & Freeman, R. (2022). Past, present, and future of the Living Planet Index. *BioRxiv*. https://doi.org/10.1101/2022.06.20.496803

Leung, B., Hargreaves, A. L., Greenberg, D. A., McGill, B., Dornelas, M., & Freeman, R. (2020). Clustered versus catastrophic global vertebrate declines. *Nature*, *588*, 267–271.

Li, J., & Hu, Q. (2021). Using culturomics and social media data to characterize wildlife consumption. *Conservation Biology*, *35*, 452–459.

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

Microsoft. (2022). *Cloud Computing Services*. Microsoft Azure. https://azure.microsoft.com/en-gb/

Millard, J. W., Gregory, R. D., Jones, K. E., & Freeman, R. (2021). The species awareness index as a conservation culturomics metric for public biodiversity awareness. *Conservation Biology*, *35*, 472–482.

Mittermeier, J. C., Roll, U., Matthews, T. J., & Grenyer, R. (2019). A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLoS Biology*, *17*, e3000146. https://doi.org/10.1371/journal.pbio.3000146

Monkman, G. G., Kaiser, M. J., & Hyder, K. (2018). Text and data mining of social media to map wildlife recreation activity. *Biological Conservation*, *228*, 89–99.

Open Analytics. (2022). ShinyProxy. https://shinyproxy.io/

Otsuka, R., & Yamakoshi, G. (2020). Analyzing the popularity of YouTube videos that violate mountain gorilla tourism regulations. *PLoS ONE*, *15*, e0232085. https://doi.org/10.1371/journal.pone.0232085

Phillips, B. B., Burgess, K., Willis, C., & Gaston, K. J. (2022). Monitoring public engagement with nature using Google Trends. *People and Nature*, *4*(5), 1216–1232. https://doi.org/10.1002/pan3.10381

Puurtinen, M., Elo, M., & Kotiaho, J. S. (2022). The Living Planet Index does not measure abundance. *Nature*, *601*, E14–E15. https://doi.org/10.1038/s41586-021-03708-8

Roll, U., Jarić, I., Jepson, P., daCosta-Pinto, A. L., Pinheiro, B. R., Correia, R. A., Malhado, A. C., & Ladle, R. J. (2021). COVID-19 lockdowns increase public interest in urban nature. *Frontiers in Ecology and the Environment*, *19*, 320–322.

Sbragaglia, V., Coco, S., Correia, R. A., Coll, M., & Arlinghaus, R. (2021). Analyzing publicly available videos about recreational fishing reveals key ecological and social insights: A case study about groupers in the Mediterranean Sea. *Science of The Total Environment*, *765*, 142672.

Sethi, S. S., Ewers, R. M., Jones, N. S., Signorelli, A., Picinali, L., & Orme, C. D. L. (2020). SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. *Methods in Ecology and Evolution*, *11*, 1182–1185.

Smith, L., Liang, Q., James, P., & Lin, W. (2017). Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, *10*(3), 370–380.

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., Townsend, S. E., Carbone, C., Rowcliffe, J. M., Whittington, J., Brodie, J., Royle, J. A., Switalski, A., Clevenger, A. P., Heim, N., & Rich, L. N. (2017). Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, *15*, 26–34.

Thompson, R. M., Hall, J., Morrison, C., Palmer, N. R., & Roberts, D. L. (2021). Ethics and governance for internet-based conservation science research. *Conservation Biology*, *35*, 1747–1754.

Twitter. (2022a). Twitter API for Academic Research. https://developer.twitter.com/en/products/twitter-api/academic-research

Twitter. (2022b). Developer agreement and policy. https://developer.twitter.com/en/developer-terms/agreement-and-policy

United Nations Environment Programme Convention on Biological Diversity (UNEP CBD). (2021). *Proposed monitoring framework for the post-2020 global biodiversity framework*. CBD, Montreal.

Wall, J., Wittemyer, G., Klinkenberg, B., & Douglas-Hamilton, I. (2014). Novel opportunities for wildlife conservation and research with real-time monitoring. *Ecological Applications*, *24*, 593–601.

Wearn, O. R., & Glover-Kapfer, P. (2017). Camera-trapping for conservation: A guide to best-practices. *WWF Conservation Technology Series*, *1*(1), 181.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio. (2022). *ggplot2: Create elegant data visualisations using the grammar of graphics*. R package.

Wikimedia. (2022). REST API documentation. https://wikimedia.org/api/rest_v1/

Wong, Y., & Rosindell, J. (2022). Dynamic visualisation of million-tip trees: The OneZoom project. *Methods in Ecology and Evolution*, *13*, 303–313.

How to cite this article: Johnson, T. F., Cornford, R., Dove, S., Freeman, R., & Millard, J. (2023). Achieving a real-time online monitoring system for conservation culturomics. *Conservation Biology*, e14096. https://doi.org/10.1111/cobi.14096