Realising the potential of real-time online monitoring for conservation culturomics

Thomas Frederick Johnson[1]*, Richard Cornford[2], 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, WC1E 6BT, UK
[4] Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
[5] Leverhulme Centre for Demographic Science, Nuffield College, University of Oxford, Oxford OX1 3UQ, United Kingdom

*Corresponding author(s): Joseph Millard, joseph.millard@nhm.ac.uk; Thomas Frederick Johnson, t.f.johnson@sheffield.ac.uk

Abstract
Environmental monitoring is increasingly shifting towards 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 towards a network of real-time platforms, given its existence almost exclusively in the digital realm. Here we describe a set of considerations associated with the development of real-time monitoring platforms for conservation culturomics. We then 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 (hosted here: https://joemillard.shinyapps.io/Real_time_SAI/), with plans to make the underlying data queryable via an API independent of the platform. The real-time SAI will represent the first real-time and entirely automated conservation culturomic platform, and one of the first within the discipline of ecology. We conclude by envisioning a future for real-time monitoring, presenting a general framework for real-time monitoring in ecology, and calling for an online real-time observatory that can evolve with the structure of the web.

Introduction

Real-time monitoring has revolutionised 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 like eDNA (Garlapati et al., 2019), remote sensing (Steenweg et al., 2017), acoustics (Sethi et al., 2020), animal satellite telemetry (Wall et al., 2014) and culturomics (Ladle et al., 2016), offer potential pathways towards real-time monitoring of biodiversity and ecosystem services.

Conservation culturomics (henceforth referred to as CC), a sub-field 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 its existence almost wholly in the digital realm. The study of CC has proved insightful in improving our understanding of human-nature interactions. For instance, culturomics has shown that interest in biodiversity increased during COVID-19 lockdowns (Roll et al., 2021), and that interest changes according to seasonality (Mittermeier et al., 2019). From a conservation perspective, culturomics has revealed patterns of wildlife trade (Li and Hu, 2021) and helped to gather information on wildlife-associated recreational activities (Monkman et al., 2018; Otsuka and Yamakoshi, 2020). All of the above are changing the conservationist's understanding of human-nature interactions. Unfortunately, there is often a lag between data being collected and being incorporated into published scientific outputs that can be used to inform conservation action. A real-time CC platform could address this lag, helping to create immediate actionable insights.

A prime example for the use of a real-time CC platform concerns interest or awareness in species. Tracking the spread (distribution) of invasive species is one potential application. With real-time information, species arrival could be detected through spikes in culturomic activity (e.g. Twitter posts, Wikipedia searches, citizen science records). For example, Asian Hornet Watch, a mobile app for reporting sightings of the invasive Asian Hornet in the UK, enables the rapid implementation of management to prevent further spread of an invasive species (CEH, 2017). Alert systems of this sort help to target limited conservation resources, making real-time culturomics platforms of great potential value to the field of conservation. Online CC 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.

Here we discuss a set of considerations in building a real-time CC platform, before introducing a near real-time platform for the Species Awareness Index (SAI). The SAI describes the rate of change in Wikipedia page views for ~40,000 species across 10 Wikipedia languages, adjusted for the background change in each language (Millard et al., 2021). The SAI could in theory be used to track the societal extinction of species (Jarić et al., 2022), as a complement to biological extinction measured through metrics such as the Red List Index (Butchart et al., 2007), the Living Planet Index (LPI: Collen et al., 2009), or the Biodiversity Intactness Index (De Palma et al., 2021). Our prototype platform is currently hosted online as a Shiny app (https://joemillard.shinyapps.io/Real_time_SAI/), which will update automatically each month through a batch process that runs virtually. This real-time platform is cost effective, scalable, and implementable with modest programming skills. We finish by envisioning a future for real-time monitoring in the context of conversation culturomics, and then conclude by emphasising the need for long-term joined-up thinking on CC.

Considerations for real-time conservation culturomics

There are many challenges associated with the development of real-time monitoring platforms for CC. In this section we set out four key considerations: 1) Platforms need to track robust metrics that have a meaningful interpretation in real-time; 2) Platform hosting will ideally be cost-effective and computationally efficient; 3) Platforms will need to be built such that they account for the structure of APIs, in terms of limits, permissions, and potential changes; and 4) 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. In this section, we consider each of these points in turn, with a particular focus on CC. However, many of the considerations we raise will also be valuable for any real-time monitoring program in ecology.

Robust metrics of human-nature interactions

A prerequisite for any real-time CC platform is the development of robust metrics that take some web-derived data source(s) as an input and then outputs a metric with a temporal dimension that meaningfully describes some form of human-nature interaction. Such a metric can be highly abstracted describing human-interactions in a coarse manner, or specifically oriented towards a particular population or behaviour. Highly abstracted metrics include those that set out 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. This is for two 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 is measured at an annual resolution. Some sources, such as Twitter data, are time-stamped at a very high resolution of 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. As a minimum, each metric must be at least as fast to derive as its data is collected, i.e. if data is 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 in processing the data before the derivation of any metric.

Unlike the publication of a CC 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 SAI (Millard et al., 2021), which leans heavily on the methodological backbone of the LPI. Although the LPI 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 SAI (Millard et al., 2021) or Biodiversity Engagement Indicator (Cooper et al., 2019) have not yet been sufficiently critiqued to fully understand the extent to which they are useful or meaningful. As in the LPI (Ledger et al., 2022), the authors of the SAI are receptive to feedback to ensure that the field continues to move forward. 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.

Hosting a real-time platform

A real-time platform could be hosted on any website, but this often requires specialised html programming skills which are uncommon amongst CC 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., 2022). Shiny apps are developed using R, a common programming language amongst quantitative conservationists (Lai et al., 2019), and so can be developed and maintained by CC researchers. Shiny apps are also cost-effective, as the price you pay is dependent on their use. For high-use applications, Shiny applications may struggle with high-traffic, in which case it could be worthwhile exploring services such as ShinyProxy, which are more equipped for concurrent application use (Open Analytics, 2022).

Whilst Shiny applications are suitable for hosting a real-time platform, they are ill-equipped for the intensive computation needed to derive CC metrics. A solution to this could be to use virtual machines or servers as a ‘back-end’ to derive the metric, and then 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 as they generally charge proportionate to use. By shifting all computation on to online servers, instead of local computers, there is less risk of disruption from software updates or power outages. Perhaps the biggest obstacle to using online servers is researcher’s unfamiliarity with them, as they can be dense and complex to use, but through making this shift, we can increase the resilience of platforms and improve their longevity.

Accounting for the structure of APIs

The data used in CC metrics are often drawn from application programming interfaces (APIs), which allow online hosted datasets to be queried and downloaded. These APIs are dynamic in nature, as they host a constant stream of new data. They are also subject to changing terms of use, can alter the format data is 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’s likely a program of long-term maintenance would be needed to address these issues. It's 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. Whilst API changes will need to be addressed to prevent platforms from failing, it's 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 now Twitter allows researchers access to the entire history of public tweets, with much less restrictive download limits. This presents opportunities for CC dashboards to not just monitor real-time changes, but also historical change.

Caution needs to be applied when using online culturomics datasets 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 licences for using data from websites like Wikipedia and Twitter often prevent the sharing of any raw data, and only allow the presentation of aggregated data. Finally, whilst the data are publicly available, there are still ethical considerations for using these datasets (Di Minin et al., 2021). Websites with personal profiles and data such as Twitter or Flickr should be used carefully. It’s important that any culturomics application, real-time or not, abide by the highest ethical and legal standards (Thompson et al., 2021).

Developing a funding plan for the long-term

We propose that a clear, consistent, and long-term funding plan is developed to allow real-time conservation culturomic platforms to run with longevity. A funding plan is essential to ensure sufficient personnel for the long-term development and maintenance of the platform, which will be needed as the structure of the web changes. It’s also important that funding is available to handle the cost associated with any 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 is likely not the best place for real-time platforms to be maintained. Instead, the ideal location for any real-time platform could be with a conservation or natural history organisation, that has both an established informatics team, and sits within a network of practitioners that could make use of the platform (e.g. RSPB, ZSL, Birdlife International, IUCN, UNEP-WCMC, the Natural History Museum).

A near real-time monitoring platform for the Species Awareness Index

Building a real-time version of the SAI has been a challenging endeavour. Although a prototype is now built, development will inevitably be a continual process, as bugs are overcome and new features added. In this section, we set out three core stages which were required in the building of a real-time platform for the SAI: first, we describe the core changes made to the underlying SAI code, required to enable the efficient execution of the whole pipeline in one batch process; second, we describe the Shiny app platform used to host the data outputted from our batch process; and third, we summarise the whole batch process itself. In the context of each section, we briefly highlight future areas of development. Much of the content below assumes that the reader is familiar with the SAI itself (see Millard et al 2021).

Changes to the core SAI code

In the original SAI publication, much of the code we wrote relied on the R package rLPI (Institute of Zoology, Zoological Society of London, 2022). rLPI is a package designed specifically for the calculation of the LPI, an aggregated trend representing the average change in abundance of many vertebrate populations. Given that the SAI is inspired by the LPI, it made sense at the initial stage to apply it to Wikipedia views. For a couple of reasons, we have moved away from the rLPI package, and instead have developed a set of scripts specifically for the SAI. First, to be amenable for use with the rLPI package, Wikipedia view data needed to be reformatted such that it looks like a set of population trends. Moving forward we can now use the more informative Wikipedia specific column names. Second, rLPI contains a large quantity of additional code that we did not require in the SAI. 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 also do not jack-knife the bootstrapped trend of random adjusted species pages, or yet include a form of weighting. In the original version of the code, bootstrapped trends were jack-knifed 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 jack-knife the trend by language each month, calculate some parameter that summarises 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, since 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 Supplementary information of our original paper, we also explored a set of weightings by total internet users and the number of species in which a language is represented (Millard et al 2021). Again, for both of 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 SAI is hosted online on a Shiny app (see here https://joemillard.shinyapps.io/Real_time_SAI/), which will update automatically each month through reads from an AWS droplet. We use this approach first and foremost because Shiny is a package written explicitly to integrate with R. Given that the SAI code is written in R, it therefore made sense to build the platform in Shiny, since we could just port over all the ggplot2 code (Wickham et al., 2022) used to build the visualisations 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 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).

A summary of the whole real-time SAI 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 on the fly with each instance of the app. In the case of the SAI, 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 SAI trends would render the platform unusable. Instead, therefore, we have opted to shift all core SAI 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 it's fired up. Through building the platform in this way, Shiny will always request from the most recent version of the SAI output, and it will only ever take as long to load as it does to generate the set of ggplots on the 'Trends' page. In this subsection we describe each core step of the real-time SAI pipeline that runs each month (also see Figure 1 for a schematic of each core step in the pipeline). Note that we do not describe the Shiny app itself, since this functions just to visualise the data outputted by the monthly batch process.

In the first instance, each month a Python script downloads Wikipedia views from each of the species and random pages previously included in the original SAI publication (see Millard et al 2021), and then 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 and Rosindell, 2022), and the random pages were identified using the Random Page Wikipedia API (as in Millard et al 2021). We opted not to download a different set of random pages each month for each language, since this alone would be a 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 SAI 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, 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 SAI code. This batch process consists of 7 scripts that run in sequence: 1) the derivation of individual species page view trends; 2) the derivation of individual random page view trends; 3) the bootstrapping of the random page view trend in each language; and then 4), 5), 6), and 7), which 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 Figures 1-4 on the ‘Trends’ page of the platform. These four files are stored in an AWS droplet from which the Shiny app reads each time the platform is fired up. 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’s been deployed (i.e. code running on our virtual machine can access local directories on that virtual machine, but a deployed Shiny app can’t). 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.

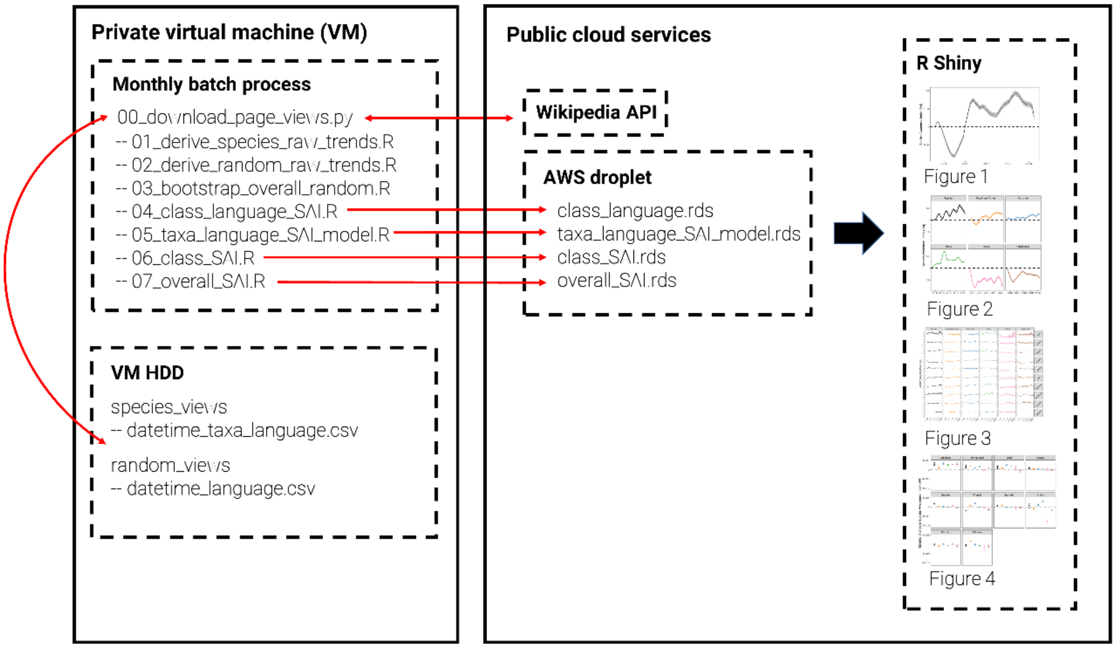

Figure 1. Schematic summarising the real-time SAI pipeline that runs on a virtual machine each month. The whole pipeline is split between software running on a virtual machine and publicly available cloud service software. Each month a batch process runs on the virtual machine, which downloads the new set of views for that month from the Wikipedia API (according to the date of the previous set), calculates the daily average views per month for both the random and species pages, and then writes those files to the virtual machine HDD (script 00). That initial Python script then runs 7 other scripts that result in four rds files, each of which are written to an AWS droplet on the cloud, and then read into Shiny whenever an instance is fired up.

Envisioning a future for real-time monitoring in ecology

In the previous section we set out how a real-time monitoring platform can be built for an index of changing awareness of species. In the future, our hope is that platforms of this sort can sit within an online observatory of similar metrics. Such a platform would help to realise the aim of many in the CC community (Jarić et al., 2020) for an observatory that tracks analogues of societal extinction (Jarić et al., 2022). Ideally, such a monitoring system would incorporate a number of datasets, 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's vital that we ensure developed metrics add value. Regardless of these problems, however, a real-time digital observatory is feasible, since the underlying infrastructure in leveraging a real-time platform will be the same irrespective of data source.

Developing a real-time observatory on CC data is a natural first-step, as an extensive array of data sources exist in the digital realm and so are amenable to the automated work-flow we present. However, this work-flow could also form the basis of a general framework for real-time monitoring in ecology, defined by four stages: 1) data is collected in real-time; 2) on a virtual machine, collected data is processed into metrics of change; 3) metrics of change are visualised with a front-end tool such as R Shiny; 4) these front-end platforms then deliver insight to policy-makers in near real-time (Figure 2). 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 CC and ecology more generally. Such a development would help to realise a core aim of the CBD (Convention on Biology Diversity), for a set of ambitious and modern indicators that are compiled and updated regularly (UNEP CBD, 2021).

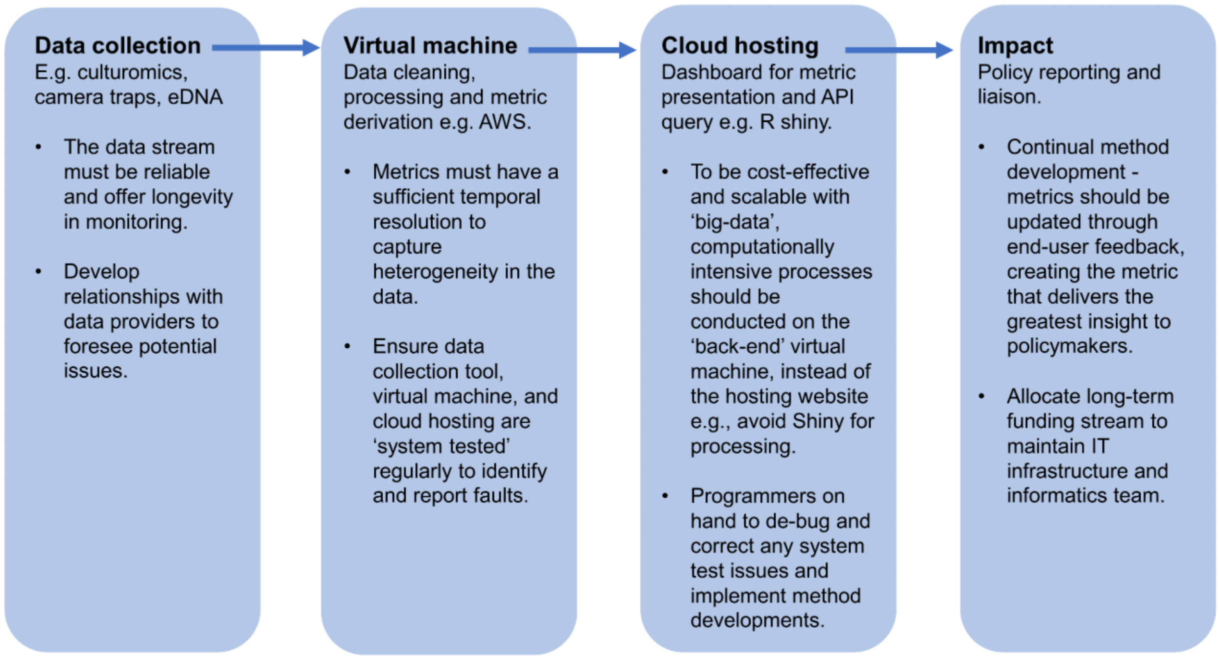

Figure 2. General framework for developing a real-time monitoring program in ecology. The four boxes represent distinct stages, moving from data collection to delivering an impactful metric. Inside each box we bullet-point considerations for each stage.

Summary

Conservation culturomics has reached a critical point at which it is ideally placed to lead the way on real-time monitoring. Here we discuss a number of key considerations in realising this potential, before introducing a near-real time platform for the SAI. Our new platform will become an ongoing project, with new features added and bug fixing on a continual basis. Such a platform demonstrates the potential of culturomics to give insights on human-nature interactions, as they play out in the physical realm. Our eventual 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, helping to realise a core aim of the culturomics community.

Acknowledgements

First and foremost, thanks to Wikipedia for providing freely accessible page view data via their APIs. Thanks to Onezoom (James Rosindell and Yan Wong; https://www.onezoom.org/) for supplying taxonomic reference data. Thanks also to the RSPB, NERC (GLiTRS, NE/V006800/1), the London NERC DTP, Andy Purvis, Richard Gregory, and Melinda Mills, who all provided support and or funding during the development of this platform.

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. https://doi.org/10.1016/j.jnc.2020.125794

AWS, 2022. Cloud Computing Services - Amazon Web Services (AWS) [WWW Document]. URL https://aws.amazon.com/ (accessed 8.8.22).

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. https://doi.org/10.13140/RG.2.2.26970.06088

CEH (Centre for Ecology and Hydrology), 2017. Asian Hornet Watch

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., shiny: Web Application Framework for R.

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. https://doi.org/10.1111/j.1523-1739.2008.01117.x

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. https://doi.org/10.1016/j.biocon.2018.12.004

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. Sci Rep 11, 20249. https://doi.org/10.1038/s41598-021-98811-1

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. https://doi.org/10.1111/cobi.13708

Docker, 2022. Home - Docker [WWW Document]. URL https://www.docker.com/ (accessed 8.8.22).

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., Link to external site, this link will open in a new window, 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. https://doi.org/10.1007/s11157-019-09501-4

Google Cloud, 2022. Cloud Computing Services [WWW Document]. Google Cloud. URL https://cloud.google.com/ (accessed 8.8.22).

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T. and Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, *34*(23), pp.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. https://doi.org/10.1093/biosci/bix025
Institute of Zoology, Zoological Society of London, 2022. rlpi package (beta).
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., Souza, A.T., Wolter, C., 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. https://doi.org/10.1016/j.tree.2021.12.011
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. https://doi.org/10.1016/j.asoc.2020.106582
Ladle, R.J., Correia, R.A., Do, Y., Joo, G.J., Malhado, A.C., Proulx, R., Roberge, J.M. and Jepson, P., 2016. Conservation culturomics. *Frontiers in Ecology and the Environment*, *14*(5), pp.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. 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. https://doi.org/10.1038/s41586-020-2920-6
Li, J., Hu, Q., 2021. Using culturomics and social media data to characterize wildlife consumption. Conservation Biology 35, 452–459. https://doi.org/10.1111/cobi.13703
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 [WWW Document]. URL https://azure.microsoft.com/en-gb/ (accessed 8.8.22).
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. https://doi.org/10.1111/cobi.13701
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. https://doi.org/10.1016/j.biocon.2018.10.010
Open Analytics, 2022. ShinyProxy [WWW Document]. URL https://shinyproxy.io/ (accessed 8.8.22).
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 n/a. 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., da Costa-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. Front Ecol Environ 19, 320–322. https://doi.org/10.1002/fee.2374

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. https://doi.org/10.1016/j.scitotenv.2020.142672

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. https://doi.org/10.1111/2041-210X.13438

Smith, L., Liang, Q., James, P. and 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), pp.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. https://doi.org/10.1002/fee.1448

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. https://doi.org/10.1111/cobi.13778

Twitter, 2022a. Twitter API for Academic Research | Products [WWW Document]. URL https://developer.twitter.com/en/products/twitter-api/academic-research (accessed 6.28.22).

Twitter, 2022b. Developer Agreement and Policy – Twitter Developers [WWW Document]. URL https://developer.twitter.com/en/developer-terms/agreement-and-policy (accessed 8.8.22).

UNEP CBD (United Nations Environment Programme Convention on Biological Diversity). 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. https://doi.org/10.1890/13-1971.1

Wearn, O.R. and Glover-Kapfer, P., 2017. Camera-trapping for conservation: a guide to best-practices. *WWF conservation technology series*, *1*(1), p.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.

Wikimedia, 2022. REST API Documentation [WWW Document]. URL https://wikimedia.org/api/rest_v1/ (accessed 8.8.22).

Wong, Y., Rosindell, J., 2022. Dynamic visualisation of million-tip trees: The OneZoom project. Methods Ecol Evol 13, 303–313. https://doi.org/10.1111/2041-210X.13766