PLOS SUSTAINABILITY AND TRANSFORMATION

RESEARCH ARTICLE

Tracking worldwide interest in sustainable development goals using culturomics

Ricardo A. Correia[1,2,3]*, **Enrico Di Minin**[1,2,4]

1 Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland, **2** Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland, **3** Biodiversity Unit, University of Turku, Turku, Finland, **4** School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

* raheco@utu.fi

OPEN ACCESS

Citation: Correia RA, Di Minin E (2023) Tracking worldwide interest in sustainable development goals using culturomics. PLOS Sustain Transform 2(6): e0000070. https://doi.org/10.1371/journal.pstr.0000070

Editor: Tien Ming Lee, Sun Yat-Sen University, CHINA

Received: November 2, 2022

Accepted: June 5, 2023

Published: June 30, 2023

Data Availability Statement: Data used in this study are publicly available at https://trends.google.com/trends

Funding: RAC acknowledges financial support from the KONE Foundation (#202101976) and the Academy of Finland (#348352). EDM thanks the European Research Council (ERC) for funding under the European Union's Horizon 2020 research and innovation program (grant agreement 802933). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The lack of robust and timely data continues to be a limiting factor in the capacity to monitor progress towards the United Nations Sustainable Development Goals (SDGs). Developing indicators based on big data has been proposed as a suitable approach to overcome this challenge but such developments have largely focused on earth observation data. Digital data representing aspects of human behaviour, such as information-seeking behaviour, hold great potential to monitor interest and engagement with sustainability topics. Using worldwide data from online searches carried out through Google's search engine, we explore how interest in the seventeen SDGs has changed over time since the goals were first proposed and assess how the COVID-19 pandemic affected these dynamics. Our results suggest a growing trend for searches related to the SDGs since they came into place in 2016. Interest was mostly directed at social and economic-related goals in the first years, but increasing interest in environmental goals has greatly increased in the last years. The onset of the COVID-19 pandemic caused significant disruptions in internet searches for the SDGS, but interest has recovered since then and continues to grow. Searches for the different SDGs are also more frequently connected following the pandemic, possibly indicating a growing awareness of the inter-related nature of sustainability goals. While further progress is needed to captivate broader engagement with the SDGs, our analysis suggests some progress has been achieved and highlights the potential of digital data to improve our understanding of public interest and engagement with sustainability topics. To materialize this potential, we outline specific areas where digital data can enhance capacity to monitor and shape sustainability progress.

Author summary

Monitoring progress towards the United Nations Sustainable Development Goals is essential to ensure they are achieved by the 2030 deadline. However, data available for this purpose is scarce and often not available in near real-time which prevents prompt action. Digital big data have been proposed as a potentially source of useful and timely information for monitoring progress towards the Sustainable Development Goals, but

Competing interests: The authors have declared that no competing interests exist.

development of suitable indicators and validation of their application has lagged. We propose that the volume of internet searches can be used as an indicator for monitoring engagement with the Sustainable Development Goals. Using data from internet searches carried out through Google's Search Engine, we show that interest in the different Sustainable Development Goals has continued to increase since they were officially announced, and despite a period of lower interest caused by the COVID-19 pandemic. We also suggest that interest in the different goals has become more connected which can indicate broader awareness of the interconnected nature of sustainability goals. Our study is a demonstration of how digital big data can be used to develop new indicators of progress towards broader engagement with sustainable development goals and we highlight opportunities for the further development of digital indicators.

1. Introduction

The ratification of the seventeen Sustainable Development Goals (SDGs) that comprise the United Nations' 2030 Agenda for Sustainable Development in 2015 represented a landmark for sustainable development efforts [1]. The SDGs constitute commitments to enact progress towards global sustainability through a set of integrated goals (e.g., ending poverty and famine go together with tackling climate change and conserving nature) that aim to balance the economic, social, and environmental dimensions of sustainable development. The roadmap represented by the SDGs seeks to engage people from multiple regions, cultures, and sectors of society, as involvement of all these groups is essential for the agenda to be successful and its achievements irreversible [1].

Monitoring relevant progress towards the targets defined in the SDGs [2] is crucial to ensure the success of the 2030 Agenda for Sustainable Development. Difficulties in assessing progress towards the Millennium Development Goals (precursors to the SDGs) associated with data availability, accuracy and timeliness [3], ensured that the identification of adequate data and indicators for monitoring progress towards the SDGs was priority since their establishment [1]. Big data has emerged as a suitable source of information for monitoring the SDGs [4,5]. The use of large digital data sources and methods is especially attractive in the face of growing demands for timely data, increasing costs of field data collection, and dwindling budgets [6]. A review of advances in the use of big data for monitoring the SGDs highlighted that datasets and methods have been explored for fifteen out of the seventeen SDGs, covering over fifty specific sustainable development targets through more than sixty indicators [7]. However, the same study also highlights that many SDGs are addressed by only one or two indicators, and that most of these indicators draw from satellite and earth observation data whereas behavioural and opinion data remain comparatively unexplored. This contrasts with the increasing uptake of digital datasets and methods in areas affiliated with sustainability goals.

For instance, the emerging research area of conservation culturomics has been spurred by the expanding use of large digital datasets and approaches for biodiversity conservation [8]. Culturomics methods (see Box 1) allow researchers to analyse the content available online through digital platforms, for example in the form of text, images, and videos, and how people engage with such content by searching for, visualizing, assessing, and sharing it (Fig 1). The novelty of these methods implies that their full potential in relation to the SDGs is still to be explored, but exciting applications are emerging. Data from internet searches, for example, have been shown to predict well a range of real-world dynamics associated with employment growth [9], medical assistance seeking [10,11], and consumer behaviour [12], all of which have

Box 1 – What is culturomics?

Culturomics can be broadly defined as the quantitative study of human culture using large datasets. The approach was first proposed by Michel and colleagues [54] who defined it as "the application of high-throughput data collection and analysis to the study of human culture". The authors argue that culturomics methods can be seen as analogous to the microscope, allowing scientists to explore and reveal aspects of cultural dynamics that would go amiss otherwise.

In their pivotal study, Michel et al. [54] used a corpus of millions of digitized books available through Google Ngram Viewer to explore cultural dynamics between 1800 and 2000. They show, for example, how the content of books can be used to study how people rise to fame and detect censorship and idea suppression. Despite the focus of the original analysis on the textual content of books, the authors recognized the need to expand culturomics analyses to other cultural products such as newspapers, maps, and artworks, and therefore to expand culturomics methods also to non-textual information. Similar views have also been expressed by other authors [55], and there is increasing recognition that the reach of culturomics methods can extend beyond texts supported by ongoing methodological developments in image and video analysis [50].

Since the original study was published, culturomics methods have been adopted in multiple areas of scientific inquiry, including health [56], psychology [57], economics [58], and biodiversity conservation [8]. While culturomics methods are not without limitations [50], they provide a complementary approach to other methodologies for the study of the dynamics of human culture. The reach and scope of culturomics methods could therefore represent a worthy addition to the toolbox of researchers working on sustainability topics.

clear links to sustainability goals. Long-term trends in growing (or dwindling) information-seeking behaviour can be indicative of public (dis)engagement [13–15], whereas shorter-term dynamics may provide clues to how specific events may affect sustainability progress. The onset of the recent COVID-19 pandemic provides a good case in point. While the emergence of the COVID-19 pandemic has been linked to unsustainable behaviours driving the ongoing biodiversity and sustainability crises [16], and there is evidence that the pandemic disrupted interest in topics related to health, tourism, and the environment [17,18], its consequences on sustainability efforts remain uncertain. Some scientists have outlined the need to rethink the SDGs in the wake of the pandemic [19], while others argue the pandemic re-emphasizes the importance of the SDGs as catalyst for progress [20]. Either way, continued monitoring of societal interest in sustainability topics will be paramount to ensure progress towards sustainability.

Here, we use a culturomics approach to track worldwide interest in the SDGs since their establishment at the beginning of 2016 and through the Covid-19 pandemic to the end of 2021. Firstly, we assess the broad temporal trends in search interest towards the main dimensions of sustainability (economic, environmental, and social). We then characterize the temporal dynamics of search interest for each of the seventeen SDGs individually and test whether the onset of the pandemic caused any significant disruption on search interest. We also check the relationship between online searches for different goals and assess whether their

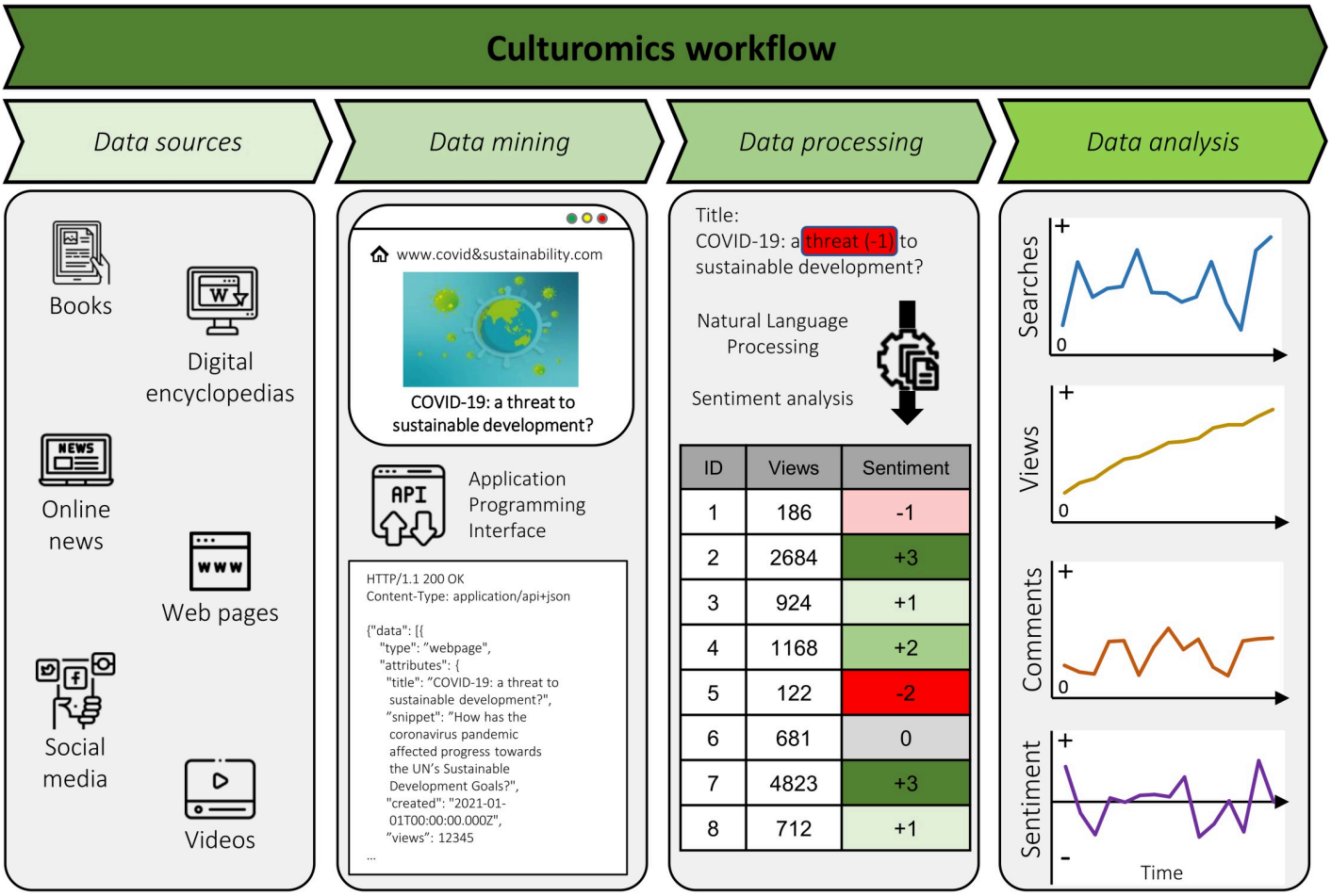

Fig 1. Schematic framework describing the process of a) data sourcing, b) data mining, c) data processing, and d) data analysis involved in culturomics analyses.

https://doi.org/10.1371/journal.pstr.0000070.g001

association increased or decreased after the pandemic to explore how perceptions of the interrelated nature of the SDGs were impacted by the pandemic.

2. Methods

We carried out a culturomics analysis of search interest using time-series data of weekly search volumes for each Sustainable Development Goal (SDG) based on information from internet searches carried out on Google's search engine between January 2016 and December 2021. We used the times-series data to i) characterize temporal trends in search volume for each sustainability dimension and each individual SDG, ii) investigate breakpoints in search volume trends for each individual SDG, and iii) assess differences in the temporal correlation of search volume between SDGs before and after the Covid-19 pandemic. We grouped SDGs under economic, environmental or social dimensions (see Table A in S1 Text) depending on which dimension the goal most closely relates to, following the categorization provided by Zheng and colleagues [21]. All analyses were implemented using R software version 4.1.3 [22] and figures were prepared using package ggplot2 for R [23].

2.1 Search volume data

We obtained time-series data of relative search volume for each SDG based on internet searches performed on Google's search engine. Specifically, we obtained weekly relative search volume data for the period between January 1 2016 and December 31 2021 (n = 313 weeks) from the Google Trends (see https://trends.google.com/) Application Programming Interface. We limited our sampling period to the end of 2021 given that changes to the Google Trends algorithm were implemented on January 1st, 2022.

Google Trends returns relative search volume data ranging from 0 to 100; the maximum value is assigned to the highest proportion of total searches observed during any given week of the sampled period and all other values are scaled in relation to it. However, because Google Trends returns search volume data that is scaled relative to the most searched for topic, and because the number of topics allowed in a single search is limited, an iterative data collection process was necessary to ensure search volume was comparable between goals [24]. Specifically, we first identified the Google Knowledge Graph entities pertaining to each SDG (see Table A in S1 Text), which were used to perform topic searches in Google Trends. Topic searches based on Knowledge Graph entities include searches for groups of terms that share the same concept in any language, thus accounting for language differences (see https://support.google.com/trends/answer/4359550).

We then carried out multiple topic searches for groups of up to five SDG topics (the maximum number of topics allowed in a single search), ensuring one common topic between each search (i.e., first search including SDGs 1,2,3,4,5, second search including SDGs 1,6,7,8,9, and so on). The values returned for the common SDG topic in either search can then be used to estimate a scaling factor between searches, which we used to rescale all search values to a single common scale. Following previous approaches [24–26], the scaling factor was calculated as the coefficient of a linear regression between the weekly values of either search. The SDGs used to calculate the scaling factor were selected iteratively because the scaling factor cannot be accurately calculated when zero values are present in the data. Therefore, the SDG used for scaling was selected for each search pair based on i) the highest number of non-zero values between searches and ii) a regression R2 value above 0.95 (which indicates a strong correlation between the values of each search). The weekly values of search volume for each SDGs were then rescaled using the calculated coefficient to ensure estimates were comparable between SDGs. The resulting data provides an estimate of the search volume with which each SDG was searched for every week over the sample period (6 years) relative to the other goals. This metric can be seen as a proxy for the underlying degree of public interest in each SDG over the sampled period. All time-series data were handled using package tsibble [27]. This software package was selected because it implements tidy data principles [28] (i.e., it provides a standardized way to organize information in a dataset), a feature that is not available in alternative packages for the same software.

2.2 Time-series analysis

We used the time-series of relative search volume data pertaining to each SDG to calculate the search volume associated with each sustainability dimension by adding the search volume of each individual SDG most associated with that dimension. These values were rescaled from 0–100 following the same rescaling process used on the original data and used to calculate the proportion of search interest in each week for the three sustainability dimensions [29]. Then, for each individual SDG, we also calculated i) the average relative search volume for each SDG over the sampling period, ii) the temporal trend of relative search volume for each SDG, and iii) the existence of relevant breakpoints in the time series data for each SDG. We calculated

the average weekly relative search volume using function 'mean' from R package *base* and used Mann-Kendall tests to detect trends in the weekly time series data for the 6 years sampled (see Table B in S1 Text). Mann-Kendall tests were implemented using the 'MannKendall' function in package Kendall [30]. The Mann-Kendall test is a robust non-parametric test to detect trends in time series analysis when data are not normally distributed [31]. The Mann-Kendall test calculates Kendall's tau statistic (τ) [32] using time-series data and tests this statistic against a null hypothesis of no trend [33]. A positive or negative tau value indicates an upward or downward trend respectively. Mann-Kendall tests have been previously used to detect the existence of significant temporal trends in relative search volume data [34].

Finally, we tested for the existence of significant breakpoints in the time-series of relative search volume data for each SDG using function 'bfast' from package *bfast* [35]. For the detection of breakpoints in the time series, we regressed the relative search volume data against the date of sampling using the minimal number of twelve observations (i.e., weeks) in each segment to detect the breakpoint using a harmonic seasonal model [36]. This calculation returns the estimated date of detected breakpoints as well as a confidence interval for the estimated date. We considered the COVID-19 pandemic affected interest in SDGs when a breakpoint and its 95% confidence interval overlapped with March 11 2020, when the World Health Organization first characterized the Covid-19 infection as a pandemic (see https://www.who.int/news/item/29-06-2020-covidtimeline).

2.3 Temporal correlation analysis

If the vision that SDGs are to be integrated and indivisible in nature is true also in the public mind, then it might be expected that internet searches for one SDG will result in concurrent searches for associated SDGs. To assess this relationship, we calculated the temporal correlation between search volume data for each SDG pair. Specifically, we assessed the correlation between searches for SDG pairs carried out in the same time-period, as this would best reflect the degree of integration between SDGs (although we recognize that lagged associations may also exist, for example see Nghiem and colleagues [34]). Assessing the correlation between internet searches requires that time-series data are stationary, but the presence of significant temporal trends rendered our data non-stationary. Hence, prior to the correlation analysis, we used function 'STL' from package *feasts* [37] to separate the trend and seasonal components of the time series, keeping only the remainder for analysis. In summary, this function applies a sequence of data smoothing operations using a locally weighted regression (loess) smoother to extract each component of the time series (see [38] for details on the formula used for time-series decomposition). Package *feasts* provides a collection of tools for processing and analysing time-series data, and was selected for use because it is also based on tidy data principles [28] and is thus able to interact with data objects generated by package *tsibble* used in earlier steps of the work.

We then performed a two-sided Pearson's correlation test between the time-series of relative search interest of unique SDGs pairs using function 'cor.test' from package stats. We assess the correlation between search interest for each SDG pair for the whole sampled period, and for the periods before and after the Covid-19 pandemic was declared. We identified the SDG pairs for which the Pearson's correlation test identified a significant positive or negative correlation and used them to plot a network of temporal correlations between search interest in SDGs. The estimated correlation coefficients (Pearson's r) for each SDG pair and for each period assess are available in Figs A-B in S1 Text.

3. Results

Our analysis of time series data representing worldwide internet searches indicates that interest has grown for the three key dimensions of sustainability since the SDGs came into effect at the beginning of 2016 (Fig 2A). Searches for SDGs associated with the economic and social dimensions dominated the earlier years, but since then the proportion of searches have become more similar between the three dimensions (Fig 2B), suggesting that interest in SDGs linked with the environmental dimension has grown faster in the last few years. These patterns are confirmed through the detailed analysis of search interest for each SDG individually (Fig 3). An increase in search volume was found for all SDGs between January 2016 and December 2021 (Mann-Kendall trend test, $p \leq 0.001>$; Supplementary Table 2), with the trend ranging between 0.489 for SDG 16 focusing on peace, justice, and strong institutions, and 0.686 for SDG 13 related to climate action. SDGs associated with the environmental dimension tended to show slightly stronger positive trends (median = 0.606) than those associated with economic (median = 0.570) or social (median = 0.567) dimensions, but no significant differences were observed in the trends between groups (ANOVA, F-value = 0.757, p = 0.488). Mean relative search volume was highest for SDG 4 (Education) and lowest for SDG 9 (Industry, Innovation, and Infrastructure).

Importantly, we also observed disruptions to growing interest in the SDGs in association with the onset of the COVID-19 pandemic. Specifically, we identified multiple breakpoints in the time-series associated with decreasing search interest immediately before or overlapping with the declaration of the COVID-19 pandemic for seven of the seventeen SDGs (Fig 3). The affected SDGs relate to all dimensions of sustainability–economic, environmental, and social–

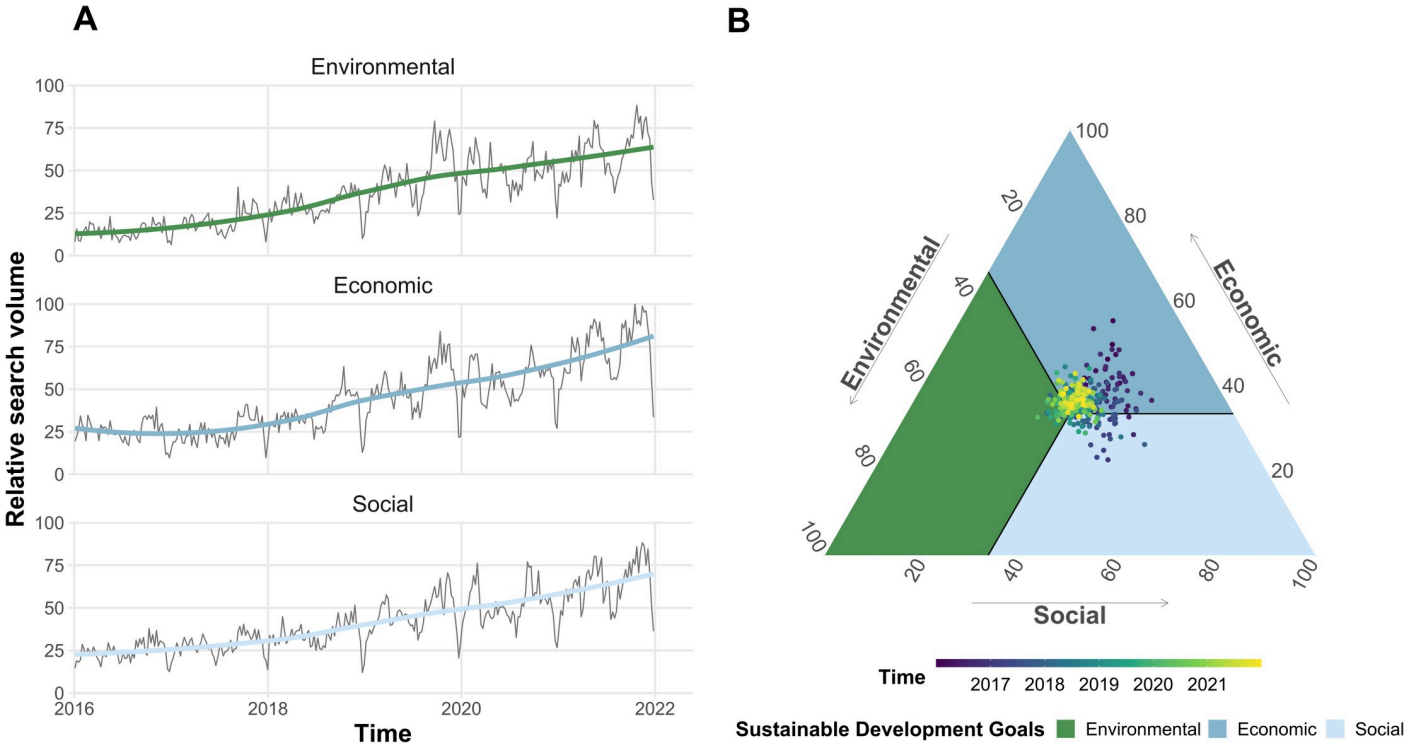

Fig 2. Summary of temporal trends in search interest towards the environmental (green), economic (dark blue) and social (light blue) dimensions of sustainable development. Temporal trends are represented in relative weekly search volume (A) and the proportion of all weekly search volume (B) for each sustainability dimension.

https://doi.org/10.1371/journal.pstr.0000070.g002

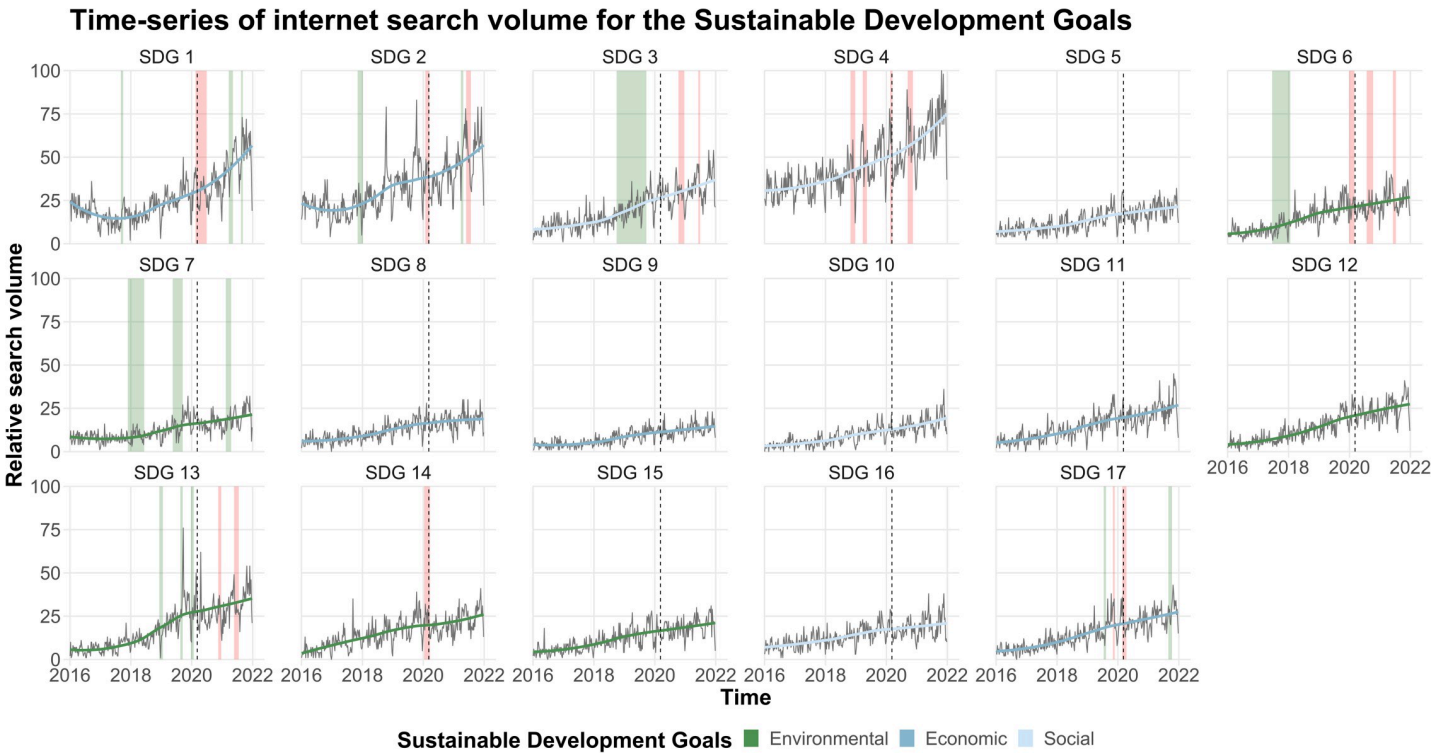

Fig 3. Temporal trends in search interest for the seventeen United Nations Sustainable Development Goals (SDGs). The time-series of relative search volume data is represented for each SDG in grey, whereas the smoothed trend is highlighted in colour for SDGs associated with environmental (green), economic (dark blue) and social (light blue) dimensions of sustainable development. Vertical dashed bars represent the week when the World Health Organization first declared the coronavirus (SARS-CoV-2) outbreak as a pandemic. Vertical coloured bars indicate periods (95% confidence interval) where a positive (green) or negative (red) breakpoint was observed in the search volume time series.

https://doi.org/10.1371/journal.pstr.0000070.g003

but were more commonly associated with goals related to economic (SDG 1—Poverty, SDG 2—Hunger, and SDG 17—Partnerships between goals), and environmental (SDG 6 –Water, SDG 13 –Climate, and SDG 14 –Marine life) dimensions than social ones (SDG 4 –Education). The disruptions resulted in marked decreases in search volume and momentaneous negative trends, suggesting the pandemic may have shifted public attention patterns. Internet searches related to the other ten SDGs continued a growing trend and no significant breakpoints were identified in relation to the pandemic.

We also found that searches for SDGs were only weakly associated in time before the pandemic, but the strength of their association increased after the onset of the COVID-19 pandemic (Fig 4). The temporal correlation in search interest between SDG pairs (i.e., the degree with which a high or low volume of searches for one SDG correspond to a similar volume in another goal) for the period between January 2016 and February 2020 was significantly lower than that observed in the period between March 2020 and December 2021 (Paired *t*-test, $t = -9.7465$, $p \leq 0.001$). The temporal correlation between SDGs increased from an average of 0.16 for the period before the onset of the pandemic to an average of 0.26 in the subsequent period. Over one hundred SDG pairs showed an increase in the search volume correlation strength after the COVID-19 pandemic, whereas only 29 pairs saw searches become more disconnected after the COVID-19 pandemic. While some SDGs showed negative temporal correlations (i.e., growing search volume for one SDG coincided with a decrease for the other SDG, and vice-versa) between SDGs before the pandemic, which can be associated with competition for public attention, no negative correlations were observed after the COVID-19 pandemic, which

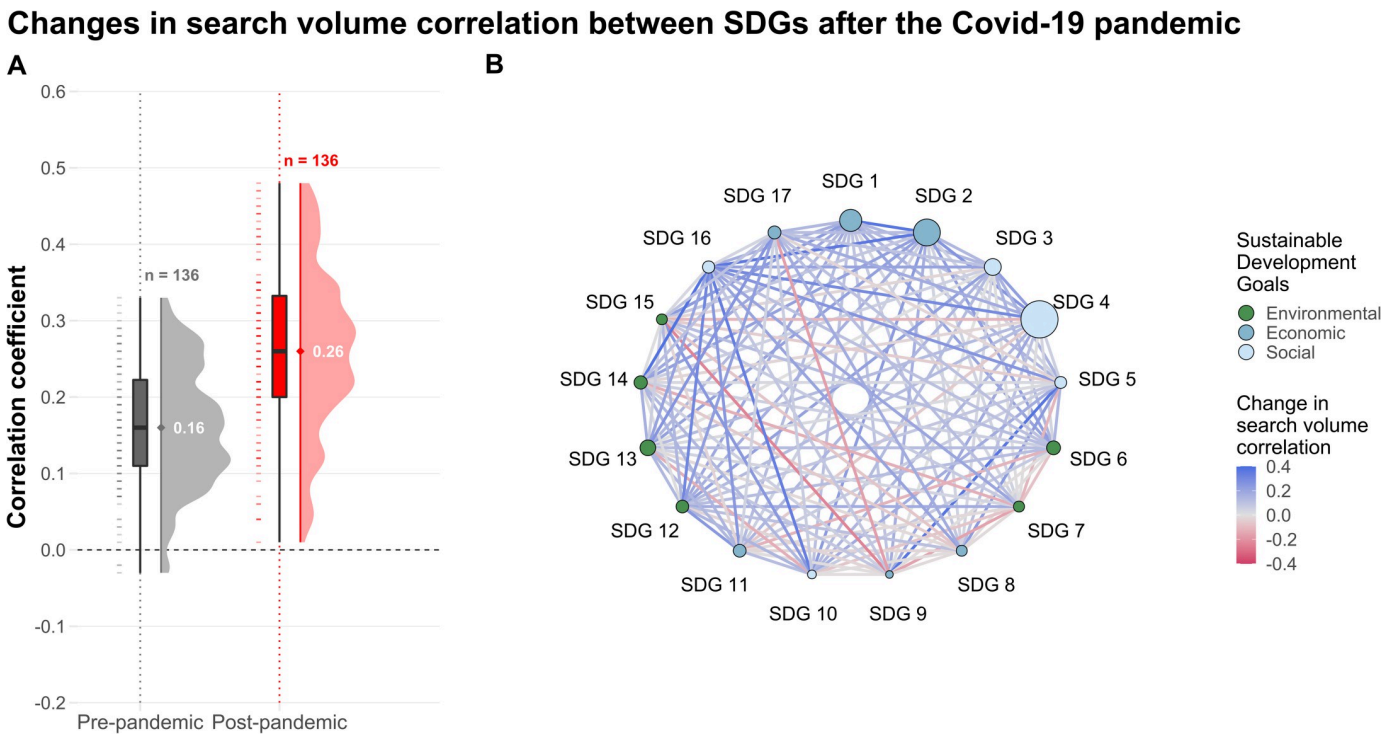

Fig 4. Summary of temporal correlations in relative search volume between United Nations Sustainable Development Goals (SDGs) and how they shifted following the coronavirus (COVID-19) pandemic. Correlation coefficients between SDG pairs are presented for the periods before (black) and after (red) the onset of the COVID-19 pandemic (A) and as the change in correlation coefficient (red representing decreasing correlation, blue representing increasing) between periods (B).

https://doi.org/10.1371/journal.pstr.0000070.g004

may be explained by a more integrated view of the SDGS and better synergy between interest in the different goals.

4. Discussion

Overall, our results indicate that internet searches for the SDGs have increased worldwide since the goals were first announced. Search interest increased for all SDGs despite the disruptions caused by the COVID-19 pandemic. Several studies have documented changes in public interest and attention towards several topics related to sustainability goals following the onset of COVID-19 pandemic and the implementation of lockdown measures to control it. For example, there were early reports of a shift in news media coverage away from the topic of climate change due to the pandemic, pointing towards a broader societal attention shift away from SDG 13 (climate action) [39]. There is also evidence that the COVID-19 pandemic shifted attention away from other prevalent diseases [40], with potential long-term consequences for health and wellbeing. At the beginning of the pandemic, it was unclear how these disruptions would affect broader societal engagement with the SDGs into the future. The sharp shift in attention away from the SDGs could have been interpreted as a sign of broader societal disengagement with the goals, supporting calls for the need to rethink and restructure the SDGs in the wake of the pandemic [19]. However, the subsequent recovery, further growth and increasing association in interest between the goals points towards the opposite direction, and suggest the SDGs are more relevant than ever [20]. In fact, many of these patterns can be interpreted considering the impacts of the COVID-19 pandemic across the world. For example, interest in SDG 4 (quality learning) has increased further and peaked after the pandemic

and is currently the goal gathering most interest (Fig 3). The pandemic disrupted education systems all over the world with the closure of schools, and evidence suggests the shift from in-person education towards remote learning is likely to have severely impacted student learning experiences [41,42]. Similarly, online interest has grown towards SDG 1 (no poverty) and SDG 3 (good health and well-being) and has become more closely aligned following the pandemic (Figs 3 and 4). The linked negative impacts of the pandemic on work and food security have been well documented in many countries, including Brazil [43], Bangladesh [44], India [45], and South Africa [46], exposing the manifold consequences of the pandemic on people's livelihoods. Taken together, our results support the idea that the SDGs are gathering interest and momentum.

Indeed, the multiple ramifications of the COVID-19 pandemic may have spurred a deeper awareness of the integrated nature of the SDGs. It has been suggested that the interrelated nature of SDGs has not been fully translated into sustainability policy and targets [47,48]. Our results suggest that awareness of their connections is weakly reflected in internet search behaviour. However, our results also suggest awareness is improving, as manifested by the more balanced distribution of search interest between economic, environmental, and social topics, and growing connections between search interest for the multiple SDGs following the pandemic. The realization that the emergence of the COVID-19 pandemic was driven by multiple economic, social, and environmental drivers has reinvigorated calls for a more holistic, 'One Health' approach to planetary health [49]. The time may be ripe to also communicate more clearly an integrated, 'One Sustainability' vision that can stimulate further progress towards sustainability across its multiple dimensions.

While our results outline patterns associated with the evolution of worldwide public interest towards the SDGs and how it was seemingly disrupted by the COVID-19 pandemic, they should be interpreted with caution. Aspects related to the nature of the data, who contributes to it and what it represents require careful consideration [50]. Indeed, we recognize that our analysis provides only a broad overview of how interest in the SDGs has progressed across the world and is unlikely to capture relevant information about these trends for many regions and countries. For example, it is well known that Google's Search Engine is not widely used in countries such as China or Russia. Therefore, our results are unlikely to accurately represent search interest in these countries. The proportion of the population with internet access also varies greatly between countries and for many countries with restricted access the internet is unlikely to accurately represent interest and awareness of the SDGs in the broader population. Hence, our results can be considered as being more representative of, and more influenced by, patterns of interest towards the SDGs in countries where internet access is widespread.

We also note that patterns of interest reported for each SDG may not fully represent search interest for the sustainability aspect they reflect. Our approach of using topic searches (see Methods) can capture multiple searches seeking information related to each SDG, for example using different terms or languages, but is unlikely to capture all relevant searches associated with each goal. A more detailed assessment of interest in the different sustainability topics using search engine data can be implemented through the identification and inclusion of other relevant topics associated with each goal (e.g., 'climate change' and 'climate emergency' in relation to SDG 13). Finally, we emphasize that the worldwide assessment we carried out here omits other important spatial, temporal and cultural dynamics that provide vital insights to ensure further progress [21]. Future studies will benefit from interdisciplinary approaches that articulate information obtained from social surveys and other more well-established methods with data obtained from digital sources to validate the results and develop more detailed, nuanced, and robust assessments that consider these dimensions.

5. Conclusions

Overall, our study demonstrates how digital approaches can help assess progress towards higher engagement with and better integrated approaches to sustainability. Relevant indicators of progress towards the SDGs are clearly needed [2] and mostly lacking for behavioural dimensions [7]. Further development of 'sustainability culturomics' and other digital approaches [5], taking advantage of internet searches, social media, and other digital data sources, can help ensure a complementary suite of indicators exists to monitor public engagement with sustainability topics at multiple scales. Such indicators can assist in assessing communication efforts by identifying ineffective or harmful framings and tracking their impacts on target audiences [51], evaluating the benefits, impacts and trade-offs of sustainable development related projects [52,53], and the development of novel leading indicators that can prove valuable in forecasting progress towards sustainability targets. More generally, the application of culturomics approaches to sustainability topics opens exciting new opportunities to foster discussions, collaborations, and experimentation towards a brighter, more sustainable future.

Supporting information

S1 Text. Information about the topics used for data collection pertaining to the SDGs and summary statistics related to search volume for each SDG and their temporal correlation for the periods before and after the COVID-19 pandemic was declared.
(DOCX)

Author Contributions

Conceptualization: Ricardo A. Correia, Enrico Di Minin.

Data curation: Ricardo A. Correia.

Formal analysis: Ricardo A. Correia.

Funding acquisition: Ricardo A. Correia, Enrico Di Minin.

Investigation: Ricardo A. Correia.

Methodology: Ricardo A. Correia.

Supervision: Enrico Di Minin.

Validation: Ricardo A. Correia, Enrico Di Minin.

Visualization: Ricardo A. Correia.

Writing – original draft: Ricardo A. Correia.

Writing – review & editing: Ricardo A. Correia, Enrico Di Minin.

References

1. United Nations General Assembly. Resolution adopted by the General Assembly on 25 September 2015. United Nations; 2015. Available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenElement.
2. Hák T, Janoušková S, Moldan B. Sustainable Development Goals: A need for relevant indicators. Ecological Indicators. 2016; 60: 565–573. https://doi.org/10.1016/j.ecolind.2015.08.003
3. Sachs JD. From Millennium Development Goals to Sustainable Development Goals. The Lancet. 2012; 379: 2206–2211. https://doi.org/10.1016/S0140-6736(12)60685-0 PMID: 22682467
4. United Nations Statistical Commission. Big data and modernization of statistical systems. New York: United Nations; 2014.

5. Ghermandi A, Langemeyer J, Van Berkel D, Calcagni F, Depietri Y, Egarter Vigl L, et al. Social media data for environmental sustainability: A critical review of opportunities, threats, and ethical use. One Earth. 2023; 6: 236–250. https://doi.org/10.1016/j.oneear.2023.02.008
6. Tam S-M, Van Halderen G. The five V's, seven virtues and ten rules of big data engagement for official statistics. Statistical Journal of the IAOS. 2020; 36: 423–433. https://doi.org/10.3233/SJI-190595
7. Allen C, Smith M, Rabiee M, Dahmm H. A review of scientific advancements in datasets derived from big data for monitoring the Sustainable Development Goals. Sustain Sci. 2021 [cited 28 Jun 2021]. https://doi.org/10.1007/s11625-021-00982-3
8. Ladle RJ, Correia RA, Do Y, Joo G-J, Malhado AC, Proulx R, et al. Conservation culturomics. Frontiers in Ecology and the Environment. 2016; 14: 269–275. https://doi.org/10.1002/fee.1260
9. Borup D, Schütte ECM. In Search of a Job: Forecasting Employment Growth Using Google Trends. Journal of Business & Economic Statistics. 2022; 40: 186–200. https://doi.org/10.1080/07350015.2020.1791133
10. Young SD, Zheng K, Chu LF, Humphreys K. Internet searches for opioids predict future emergency department heroin admissions. Drug and Alcohol Dependence. 2018; 190: 166–169. https://doi.org/10.1016/j.drugalcdep.2018.05.009 PMID: 30036853
11. Adams QH, Sun Y, Sun S, Wellenius GA. Internet searches and heat-related emergency department visits in the United States. Sci Rep. 2022; 12: 9031. https://doi.org/10.1038/s41598-022-13168-3 PMID: 35641815
12. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. Proceedings of the National Academy of Sciences. 2010; 107: 17486–17490. https://doi.org/10.1073/pnas.1005962107 PMID: 20876140
13. Mccallum ML, Bury GW. Google search patterns suggest declining interest in the environment. Biodivers Conserv. 2013; 22: 1355–1367. https://doi.org/10.1007/s10531-013-0476-6
14. Funk SM, Rusowsky D. The importance of cultural knowledge and scale for analysing internet search data as a proxy for public interest toward the environment. Biodivers Conserv. 2014; 23: 3101–3112. https://doi.org/10.1007/s10531-014-0767-6
15. Caetano GH de O, Vardi R, Jarić I, Correia RA, Roll U, Veríssimo D. Evaluating global interest in biodiversity and conservation. Conservation Biology. 2023;n/a. https://doi.org/10.1111/cobi.14100 PMID: 37070377
16. Arora NK, Mishra J. COVID-19 and importance of environmental sustainability. Environmental Sustainability. 2020; 3: 117–119. https://doi.org/10.1007/s42398-020-00107-z
17. Vijay V, Field CR, Gollnow F, Jones KK. Using internet search data to understand information seeking behavior for health and conservation topics during the COVID-19 pandemic. Biological Conservation. 2021; 257: 109078. https://doi.org/10.1016/j.biocon.2021.109078 PMID: 34584274
18. Souza CN, Rodrigues AC, Correia RA, Normande IC, Costa HCM, Guedes-Santos J, et al. No visit, no interest: How COVID-19 has affected public interest in world's national parks. Biological Conservation. 2021; 256: 109015. https://doi.org/10.1016/j.biocon.2021.109015
19. Naidoo R, Fisher B. Reset Sustainable Development Goals for a pandemic world. Nature. 2020; 583: 198–201. https://doi.org/10.1038/d41586-020-01999-x PMID: 32632244
20. Ottersen OP, Engebretsen E. COVID-19 puts the Sustainable Development Goals center stage. Nature Medicine. 2020; 26: 1672–1673. https://doi.org/10.1038/s41591-020-1094-y PMID: 33037423
21. Zheng X, Wang R, Hoekstra AY, Krol MS, Zhang Y, Guo K, et al. Consideration of culture is vital if we are to achieve the Sustainable Development Goals. One Earth. 2021; 4: 307–319. https://doi.org/10.1016/j.oneear.2021.01.012
22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2022. Available from: https://www.R-project.org/.
23. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
24. Davies T, Cowley A, Bennie J, Leyshon C, Inger R, Carter H, et al. Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment. PLOS ONE. 2018; 13: e0203694. https://doi.org/10.1371/journal.pone.0203694 PMID: 30256838
25. Mammola S, Riccardi N, Prié V, Correia R, Cardoso P, Lopes-Lima M, et al. Towards a taxonomically unbiased European Union biodiversity strategy for 2030. Proc R Soc B. 2020; 287: 20202166. https://doi.org/10.1098/rspb.2020.2166 PMID: 33290682
26. Adamo M, Sousa R, Wipf S, Correia RA, Lumia A, Mucciarelli M, et al. Dimension and impact of biases in funding for species and habitat conservation. Biological Conservation. 2022; 272: 109636. https://doi.org/10.1016/j.biocon.2022.109636

27. Wang E, Cook D, Hyndman RJ. A New Tidy Data Structure to Support Exploration and Modeling of Temporal Data. Journal of Computational and Graphical Statistics. 2020; 29: 466–478. https://doi.org/10.1080/10618600.2019.1695624
28. Wickham H. Tidy Data. Journal of Statistical Software. 2014; 59: 1–23. https://doi.org/10.18637/jss.v059.i10
29. Troumbis AY, Vasios GK. Ternary plots of public interest in cultural human-nature interactions: Google Trends-based culturomics of the mediterranean olive triptych. Journal of Rural Studies. 2021; 81: 1–8. https://doi.org/10.1016/j.jrurstud.2020.11.012
30. McLeod AI. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. 2011. Available from: https://CRANR-project.org/package=Kendall.
31. Önöz B, Bayazit M. The Power of Statistical Tests for Trend Detection. Turkish J Eng Env Sci. 2003; 27: 247–251.
32. Kendall MG. A New Measure of Rank Correlation. Biometrika. 1938; 30: 81–93. https://doi.org/10.2307/2332226
33. Mann HB. Nonparametric Tests Against Trend. Econometrica. 1945; 13: 245–259. https://doi.org/10.2307/1907187
34. Nghiem LTP, Papworth SK, Lim FKS, Carrasco LR. Analysis of the Capacity of Google Trends to Measure Interest in Conservation Topics and the Role of Online News. PLOS ONE. 2016; 11: e0152802. https://doi.org/10.1371/journal.pone.0152802 PMID: 27028399
35. Verbesselt J, Hyndman R, Newnham G, Culvenor D. Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment. 2010; 114: 106–115. https://doi.org/10.1016/j.rse.2009.08.014
36. Verbesselt J, Hyndman R, Zeileis A, Culvenor D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. Remote Sensing of Environment. 2010; 114: 2970–2980. https://doi.org/10.1016/j.rse.2010.08.003
37. O'Hara-Wild M, Hyndman R, Wang E. feasts: Feature Extraction and Statistics for Time Series. R package version 0.2.1. 2021. Available from: https://CRANR-project.org/package=feasts.
38. Cleveland RB, Cleveland WS, Terpenning I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics. 1990; 6: 3.
39. Lyytimäki J, Kangas H-L, Mervaala E, Vikström S. Muted by a Crisis? COVID-19 and the Long-Term Evolution of Climate Change Newspaper Coverage. Sustainability. 2020; 12: 8575. https://doi.org/10.3390/su12208575
40. Ng R, Chow TYJ, Yang W. News media narratives of Covid-19 across 20 countries: Early global convergence and later regional divergence. PLOS ONE. 2021; 16: e0256358. https://doi.org/10.1371/journal.pone.0256358 PMID: 34469446
41. Daniel SJ. Education and the COVID-19 pandemic. Prospects. 2020; 49: 91–96. https://doi.org/10.1007/s11125-020-09464-3 PMID: 32313309
42. Donnelly R, Patrinos HA. Learning loss during Covid-19: An early systematic review. Prospects. 2021 [cited 26 Sep 2022]. https://doi.org/10.1007/s11125-021-09582-6 PMID: 34785823
43. Neves JA, Machado ML, Oliveira LD de A, Moreno YMF, Medeiros MAT de, Vasconcelos F de AG de. Unemployment, poverty, and hunger in Brazil in Covid-19 pandemic times. Rev Nutr. 2021; 34. https://doi.org/10.1590/1678-9865202134e200170
44. Shammi M, Bodrud-Doza Md, Towfiqul Islam ARMd, Rahman MdM. COVID-19 pandemic, socioeconomic crisis and human stress in resource-limited settings: A case from Bangladesh. Heliyon. 2020; 6: e04063. https://doi.org/10.1016/j.heliyon.2020.e04063 PMID: 32462098
45. Mishra K, Rampal J. The COVID-19 pandemic and food insecurity: A viewpoint on India. World Development. 2020; 135: 105068. https://doi.org/10.1016/j.worlddev.2020.105068
46. Visagie J, Turok I. Rural–urban inequalities amplified by COVID-19: evidence from South Africa. Area Development and Policy. 2021; 6: 50–62. https://doi.org/10.1080/23792949.2020.1851143
47. McGowan PJK, Stewart GB, Long G, Grainger MJ. An imperfect vision of indivisibility in the Sustainable Development Goals. Nature Sustainability. 2019; 2: 43–45. https://doi.org/10.1038/s41893-018-0190-1
48. Blanc DL. Towards Integration at Last? The Sustainable Development Goals as a Network of Targets. Sustainable Development. 2015; 23: 176–187. https://doi.org/10.1002/sd.1582
49. Garine-Wichatitsky M de, Binot A, Morand S, Kock R, Roger F, Wilcox BA, et al. Will the COVID-19 crisis trigger a One Health coming-of-age? The Lancet Planetary Health. 2020; 4: e377–e378. https://doi.org/10.1016/S2542-5196(20)30179-0 PMID: 32918881

50. Correia RA, Ladle R, Jarić I, Malhado ACM, Mittermeier JC, Roll U, et al. Digital data sources and methods for conservation culturomics. Conservation Biology. 2021; 35: 398–411. doi:https://doi.org/10.1111/cobi.13706 PMID: 33749027

51. Ballejo F, Plaza PI, Lambertucci SA. Framing of visual content shown on popular social media may affect viewers' attitudes to threatened species. Sci Rep. 2021; 11: 13512. https://doi.org/10.1038/s41598-021-92815-7 PMID: 34188096

52. Sherren K, Parkins JR, Smit M, Holmlund M, Chen Y. Digital archives, big data and image-based culturomics for social impact assessment: Opportunities and challenges. Environmental Impact Assessment Review. 2017; 67: 23–30. https://doi.org/10.1016/j.eiar.2017.08.002

53. Chen Y, Parkins JR, Sherren K. Leveraging Social Media to Understand Younger People's Perceptions and Use of Hydroelectric Energy Landscapes. Society & Natural Resources.2019; 32: 1114–1122. https://doi.org/10.1080/08941920.2019.1587128

54. Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, Team TGB, et al. Quantitative Analysis of Culture Using Millions of Digitized Books. Science. 2011; 331: 176–182. https://doi.org/10.1126/science.1199644 PMID: 21163965

55. Ladle RJ, Jepson P, Correia RA, Malhado AC. The power and the promise of culturomics. Frontiers in Ecology and the Environment. 2017; 15: 290–291. https://doi.org/10.1002/fee.1506

56. Mandl BJ, Reis BY. The language of crisis: spatiotemporal effects of COVID-19 pandemic dynamics on health crisis communications by political leaders. npj Digit Med. 2022; 5: 1–11. https://doi.org/10.1038/s41746-021-00554-w PMID: 35013539

57. Hills TT, Proto E, Sgroi D, Seresinhe CI. Historical analysis of national subjective wellbeing using millions of digitized books. Nat Hum Behav. 2019; 3: 1271–1275. https://doi.org/10.1038/s41562-019-0750-z PMID: 31611658

58. Shiller RJ. Narrative Economics. American Economic Review. 2017; 107: 967–1004. https://doi.org/10.1257/aer.107.4.967