Research article

The physics of conservation culturomics: the mass-energy-information equivalence principle to address misrepresented controversies

Andreas Y. Troumbis [*]

Biodiversity Conservation Laboratory, Department of the Environment, University of the Aegean, Greece

ABSTRACT

The application of the mass-energy-information equivalence principle developed after the experimentally demonstrated Landauer's principle on thermodynamics, entropy, and information is an unexplored but promising path in search of objectivity and compatibility between strict physical and mathematical entities and relative human behavior in biodiversity conservation issues. Conservation culturomics is proposed as the epistemic methodology and programme to trace the evolution in cultural human-nature relationships. Historically, controversies do persist between pro- vs. non- environmental opinions and policies. The proposed combination of physics and culturomics is feasible, although complex, multileveled, and depending on a series of academic, technical, and political prerequisites. In the era of staggering information technologies, Internet use proliferation and cultural relativism, reliable information on conservation knowledge vs. often unfounded story-tellings is a sine qua non for the development of badly needed modern global conservation strategies, targets, and goals.

1. Introduction

So that a reader will have a clear understanding where s/he is heading, here is how this paper is organized and constructed. Should one examine the century-old, traditional controversy between 'nature' vs 'economy, jobs, business', s/he will quickly realize that opinions –and to some extent policies-fall into some controversial spectrum ranging from 'pro-environmentalists' to 'non-environmentalists' (or 'pro-business disciples'). Within the rather dominant 'pro-environmental' discourse nowadays where international Institutions, Governments, Conventional bodies, Corporations, Environmental NGOs (ENGOs), academic communities, or mass media take part and the increasing trend of consumers willing-to-pay for environmentally friendly products and services, things are not so idyllic or straightforward as appearing *prima fasciae.* One can identify from cynical 'green-washing' industrial strategists (recall the VW car manufacturer and its corrupted C-emissions software; e.g., Laufer 2003; Ramus and Montiel, 2016; Dahl 2010) to directors and officers of ENGOs devoted primarily to fundraising, public relations, building office installations and power games (see Terborgh 1999, p. 7–8). Greening real estate markets, power, transportation, manufacturing, agriculture and food industry, logistics, banking, or trade sectors are also fashionable, especially after the SMART 2020 ICT (Information and Communications Technologies)promoted strategy for the transition to low carbon economy (GeSI –Global e-Sustainability Initiative, 2008); on a utilitarian perspective, McKinsey and Company (2009) proposed cost curves estimating the prospective annual abatement cost (in euros per ton, roughly 40 €/tCO$_2$e) of avoided emissions of greenhouse gases vs the potential of emission abatement approaches in gigatons of CO$_2$e. Further, as Diamond (2011) put it eloquently, the equation 'non-environmentalist' = 'pro--business' is astonishingly imperfect: considerable businesspeople claim adoption and acting in honest pro-environmental ways whereas abundant numbers of average laypersons adopt non-environmentalist opinions and behavior. The latter is particularly evident in cases of the economy and land-use *transitions*, either towards protected areas schemes (e.g., Blicharska et al., 2016), energy (e.g., Bridge et al., 2013), or in delimited and small scale territories such as islands (Troumbis and Hatziantoniou 2018). Eventually, the 'pro-' vs. 'non-' environmentalist scheme yet flirts with obsoletism in the context of the staggering production of information –and digital content-on environmental and conservation issues.

Floridi (2014) introduced the concept of 'infosphere' as shaping human reality; further, since Landauer's principle (1961) demonstrating that information is physical, more recent developments on the mass-energy-information equivalence principle show that a bit of information has a finite and quantifiable mass while it stores information (e.g., Vopson 2019; 2020). In that sense, we argue that the new

* Corresponding author.
  E-mail address: atro@aegean.gr.

controversy relates to the 'big communication game' on the validity, utility, technological progress, and economy of information on environmental impacts, including carbon emissions, planetary change mitigation strategies, and biodiversity conservation. The exciting part of the mass-energy-information equivalence principle is that it calculates or measures the physical dimensions of a bit of information scientifically. For instance, in the equivalence triptych, physics equations are reliable: (1) mass-energy: $m = \frac{E}{c^2}$; (2) energy-information: $k_b T \ln(2) = mc^2$; (3) information-mass: $m = \frac{k_b T \ln(2)}{c^2}$, where k_b is the Boltzmann constant and T is the temperature at which information is stored (300 K). Given these equations, three estimations of numbers of bits N_{bits} of information accumulated M_{info} for a period of n years and of the total energy necessary to create this mass of information during the same period Q_{info} might be (Vopson 2020):

$$N_{bits}(n) = \frac{N_b}{f} \cdot ((f+1)^{n+1} - 1) \quad \text{(Section III.1, hereafter)}$$

$$Q_{info}(n^{th}) = N_b \cdot k_b T \cdot \ln(2) \cdot (f+1)^n \quad \text{(Section III.1, hereafter)}$$

$$M_{info}(n) = N_b \cdot \frac{k_b T \cdot \ln(2)}{f \cdot c^2} \cdot ((f+1)^{n+1} - 1) \quad \text{(Section III.1, hereafter)}$$

f represents % growth of digital content creation, either year-to-year or after a series of n years; f might also be expressed in a relative frequency scale.

We argue that the above equations establish perspectives to quantify the relationship between information generation and valuation of ecosystem services within the conservation culturomics framework (Ladle et al., 2016). They allow for departing from the historic confinement to a value-based controversy on choices, preferences, measures and/or principles relating to specific worldviews on human environmental impacts (e.g. Pascual et al., 2017). However, conceptual impasses and discursive stagnation, similar to traditional controversy, await within the myriads of current communication networks where issues related to the mass-energy-information equivalence propagate in a mixed audience of scientists, amateurs, pseudo-scientists, populists, and conspiracy addicts. Arguments of the form "one *Google* search is equivalent to about 0.2 g of CO_2"; "the average car driven for one kilometer produces as many greenhouse gases as a thousand *Google* searches (given that EU tailpipe emissions calls for 140 g of CO_2 per kilometer driven)"; "a typical *Google* search uses half the energy as boiling a kettle of water and produces 7 g of CO_2"; "building your search query amounts to 0.0003 kWh of energy per search, or 1 kJ. For comparison, the average adult needs about 8000 kJ a day of energy from food, so a *Google* search uses just about the same amount of energy that your body burns in ten seconds"; "a typical individual's *Google* use for an entire year would produce about the same amount of CO_2 as just a single load of washing", and so on, abound in Floridi's infosphere. The new 'information' controversy is articulated around proponents of 'information-based de-materialization' of activities, and those considering ICT promoted improvements of energy efficiency are exaggerated and unwarranted and that continuation of modern lifestyles is desirable and possible.

We propose an analysis capable of comparing estimated mass and energy of digital information on conservation issues, public interest in them, and estimations of biodiversity, productivity, and global ecosystems value to escape from this entrapment. The accuracy of compared data differs since uncertainties regarding physical and biotic world metrics are non-negligible. In short, we examine whether it is possible to establish a scope able to observe physical objects along with historical and behavioral changes. We adopt the epistemological framework of conservation culturomics (Ladle et al., 2016; Sutherland et al., 2018), keeping in mind unsolved yet technical and epistemological issues (Troumbis 2019; Troumbis and Iosifidis 2020). Although much of conservation culturomics literature pays attention to the basic idea that Web –mostly *Google*-crowd search volumes on environmental and biodiversity conservation terms or topics offer near-real-time metrics of public interest in them, the full expansion of this epistemology is constituted by four additional areas engaging (1) recognition of conservation-oriented constituencies; (2) identification of emblems; (3) assessment of cultural impacts; and (4) promotion of public understanding (Ladle et al., 2016, p. 269).

The precept from the ever-growing literature on conservation culturomics is that the wealth of online "big"-data, the analysis and use of these digital resources, and computational linguistics and lexicology are or should be harnessed to understand social trends and culture-behavior changes to gain insight into broad-scale patterns of human-nature interactions and perceptions (Di Minin et al., 2015; Correia et al. 2016, 2017; Roll et al., 2016; Cooper et al., 2019; Ladle et al., 2019; Mittermeier et al., 2019; Toivonen et al., 2019). Although trends in public interest in conservation issues in general (e.g., Proulx et al., 2013; Troumbis 2017a, b; Burivalova et al., 2018; Legagneux et al., 2018; Correia et al., 2019; Troumbis 2019), through *Google Trends*-based culturomics form a consistent corpus of research, the whole big data conservation culturomics operation is instead "messy"; as Aiden and Michel (2013, p. 19), the fathers of Culturomics, have noticed a "…*typical bigdata dataset is a miscellany of facts and measurements…riddled with errors and marred by numerous, frustrating gaps… missing pieces of information… because big datasets are frequently created by aggregating a vast number of smaller datasets… some more reliable than others and each one subject to its own idiosyncrasies…*".

In this perspectives paper, analysis targets both Michel et al. (2011) and Aiden and Michel (2013) predictions on cultural evolution represented by the evolution of frequencies in use of terms (words) in digitized books; and, Vopson (2019; 2020) mass-energy-information equivalence principle calculations regarding the conservation-related information created and stored in comparison to the annual rate of total digital bits production in a n years period (see Table 1). There are two issues of primary interest: (1) the feasibility and limits of technical calculations on the mass-energy-information equivalence principle in topics related to conservation; (2) the application of calculated values of digital information mass and energy required as a simile "metric" of the relative importance of conservation campaigns in various biomes on Earth.

In practice, the analysis focuses on (1) the estimation of the total number of bits of information on conservation-related topics accumulated on the planet after n years of f% annual growth [$N_{bits}(n)$]; (2) the total energy necessary to create such digital information in a given n^{th} year, assuming f% year-on-year growth [$Q_{info}(n^{th})$]; and, (3) the total digital information mass accumulated on the planet after n years of f% growth [$M_{info}(n)$]. The ratio R_{bits} between conservation topics *[$N_{bits}(n)$]*, *[$Q_{info}(n^{th})$]* and *[$M_{info}(n)$]* divided by respective metrics on total digital human activities and platforms and, full range of thematic topics for the same period is by definition <1. The genuine interest of such an analysis is (1) what the order of magnitude of R_{bits} is? (2) does R_{bits} evolve in time and towards which direction? (3) are there differences in R_{bits} between specific conservation-related topics and biomes (refer to the Results section, figures, and tables therein).

The various metrics are compared to average quantities of net primary productivity (NPP), measured in gC/area (in 10^{12} m^2) of major ecosystem types/year. The same exercise is repeated with the valuation of the world's ecosystem services and natural capital, proposed in the literature, in \$ha^{-1}yr^{-1} transformed here in 10^{16} m^2, in 16 different biomes; as Costanza et al. (1997) argued nature's contributions to people are inseparable from the choices and decisions to make about ecological systems and their effects on human well-being. Since monetary estimations of ecosystem services are primarily anchored on "willingness-to-pay" techniques, the relationship between the mass of digital information accumulated [$M_{info}(n)$] flowing towards people and their behavioral change might be an essential signal in conservation decision-making and implementation practices (Section 3.2).

Information on cultural human-nature relationships meant for public consumption and sentiment is created and provided through the Web

Table 1. Synopsis of basic assumptions/constants used in the calculations of conservation-related mass-energy-information equivalence.

Entity	Quantity	Units	Remark	Data sources
1 byte	8 bits	1 character or number between 0-255		
Linear size of a bit	25×10^{-9}	m		Vopson (2020)
Area size of a bit	25	nm^2	Data storage densities > 1Tb/in	Vopson (2020)
Rate of digital data produced/day	2.5×10^{18} 2.5×10^9	Bytes Gb	IBM estimation (2012)	Zikopoulos et al., (2012)
Number of bits produced/day	2×10^{19}		IBM estimation (2012)	Zikopoulos et al., (2012)
N_b: Number of bits produced/year	7.3×10^{21}	bits	IBM estimation (2012)	Zikopoulos et al., (2012)
Q_{bit}	$\approx$18	meV	At room temperature 300 K	Landauer (1996)
m_{bit} rest mass of a digital bit (room temperature)	3.19×10^{-38}	kg		Vopson (2020)
$N_b \times m_{bit}$ total mass of all information/year	23.3×10^{-17}	kg		
Energy	1 J	6.242×10^{18} eV		SI
Number of searches/second/day /year	40 3.5 1.2	kB BnB TnB	Google (>80% market share)	Internet Live Stats
Growth rate of Google search volumes	Min = 10 Max = 64	%	Data on Google	Internet Live Stats
f annual growth of digital content creation	1% (conservative) 10–15% (realistic)	bits	Estimation: double-digit value; out-of-date digital content erased all-the-time	Vopson (2020)
n period of time	number	years		
k_b Boltzmann constant (at 300 K temperature)	$1,380649 \times 10^{-23}$	$J \cdot K^{-1}$		SI

(web pages, blogs, online news; online communication via e-mails, Skype, text messages; Facebook, Twitter or Linkedin; *Google* scans books, *Flickr* stores photos, *YouTube* or *Netflix* are streaming movies and/or documentaries with environmental content). Actual numbers of information quantities are often not disclosed by providers as is the number of users of such services. Such problems are built-in in what is generically called "citizen's science" the potential of which has been extensively reviewed in recent literature, e.g., Wehn and Almomani (2019). The weakness of such an analysis as of today springs before our eyes: while the overall digital bits production in *n* year-period is strictly calculated and comprises total information from all the above components, conservation-related information is but a short sub-sample of them, both regarding topics investigated and components of production and storage of digital information.

2. Methods

Research on conservation culturomics diverges from the "noble" Popperian falsification Method, broadly adopted by modern scientists and prestigious journals (Troumbis and Iosifidis 2020); it instead converges to "comparative methods" or "natural experiments", in J. Diamond's terminology (2011), "to compare natural situations differing with respect to the variable of interest" (p. 17). Diamond initially compared civilizations, populations, and environments through long periods in an attempt to understand and recite combinations of causes that drove them to collapse. The non-experimental and not-replicable "comparative method" in conservation culturomics presents obvious pitfalls (Troumbis and Iosifidis 2020). We proceed to address the questions announced in the Introduction section by relating output variables to input variables. In this paper, potential input variables are measurements or calculations of energy-mass-information (digital) quantities, and output variables relate to ecosystem services and economic valuation.

2.1. Methodological assumptions

Technically, we estimate the various bits quantities by spreadsheet-based calculations of Vopson equations, using values of parameters (refer to Table 1) and a selection of emblematic conservation-related

topics for which digital material volumes/time are sampled from various platforms.

To feed the energy-mass-information estimation equations (Sections 3.1), we adopted a series of "standard values" as they are published officially either by provider companies, specialized and respected web sites, publications listed in the Web of Knowledge/Science or personal communications with renown experts. Some of them are irrefutable: 1 byte is 8 bits and corresponds to 1 character or a number between 0-255; or, the size of a bit is 25×10^{-9} m. Others represent a most recent estimation or an average of a series of similar estimations: e.g. in 2012, an IBM team (Zikopoulos et al., 2012) estimated that 2.5×10^9 Gb are produced daily, leading to an annual rate of total digital bits production $N_b = 7.3 \times 10^{21}$ bits. Since 2012 is the middle year of the 2004–2020 period during which *Google Trends*-based culturomics is publically available, we use these values as "constant" for the entire period. Variations are also observed regarding the market share of web search engines, e.g. *Google* market share ranges between >80 to 92%. The annual growth rate f % of *Google* search volumes seems to stabilize between 10-15% during the last 5 years. The total digital data fingerprint is doubling every 2 years –on average-as data storage density technology improves, bandwidth increases and *Homo sapiens* migrates steadily onto the Internet: this is also an average since alternative estimations of increase range from 1.5 to 3 and the digital storage capacity is predicted to get measures in zettabytes [10^{21} bytes] by 2022 (CISCO 2016). A compilation of basic assumptions/constants is presented in Table 1.

2.2. Reservations and use of low estimates

A selection of 9 topics with an inclusive and ubiquitous character (e.g., sustainability, biodiversity, climate change) that undeniably motivated/generated broader research fields growth and the establishment of new academic communities and networks since late 80s-early 90s (Bettencourt and Kaur 2011; Troumbis and Iosifidis 2020) is used as a low approximation of conservation-related information. Such topics have been proposed and published in the early conservation culturomics period (e.g., McCallum and Bury, 2013; 2014; Kim et al., 2014; Nghiem et al., 2016).

Some emblematic topics such as "environment" are purposely excluded from the selected list since they are often associated in *Google*

crowd searches with extensions and/or specifications that are irrelevant to conservation-related search queries: e.g., work environment, family environment, or medical environment. Another topic confronted with similar misleadingly recorded data on the Web is the highly symbolic "conservation" – e.g. food conservation or mass conservation. The topic finally used here is "biological conservation" considered more appealing to the general audience than "conservation biology", which conveys a more accentuated scientific character. Therefore, the estimation of unstoppable digital content produced and stored in the last 30 years (from 1990-2020) regarding conservation and environmental issues is indicative compared to the total 7.3×10^{21} bits/year times 30 years. Table 2 summarizes topics and technological components used as a low estimation of the R_{bits} as defined here above.

Tracking conservation-related digital data statistics resemble sticking to a bottomless pit of disadvantages. For example, one should keep in mind that conservation-related digital bits content on the planet is stored on all traditional devices but also cloud data storage centers and end-points, e.g., PCs, smart-phones (e.g., Toivonen et al., 2019), and Internet of Things (IoT) devices, such as physical environment sensors, tracking devices attached on migratory species or networks of remote cameras spotifying animal moves in experimental sites. Therefore, the five technological components used in this analysis are a "Lilliputian" sub-sample of the digital data creation and storage in our field of interest. Overall, the 9 topics by 5 components ensemble resemble a stimulus for thought than an accurate signal on conservation-related digital data statistics. This issue is addressed in the Conclusions section of the paper.

This analysis is driven by the aspiration to identify patterns of some 'ecumenical' interest. However, cross -countries, -languages, -cultures, -generations, -access to the Internet, -literacy, -economy, -democracy or -freedom of press inequalities (e.g., Funk and Rusowsky 2014; Troumbis and Iosifidis 2020) and divergences from the incumbent western ecological modernization theory (e.g., Clausen and York 2008) make hard to explain results –which often are self-contradictory (refer to Discussion Section).

Finally, even over-sensitive instruments used for defining fundamental constants of the metrological SI report measurement uncertainties: e.g., the ultra-sensitive Kibble balance used for determining the Kilogram, as the one developed at NPL in the UK (Robinson and Schlamminger 2016) reports uncertainties of $\sim 10^{-9}$.

3. Results

Two parts compose this section. Indicative results on the calculations or estimations – and various graphical representations - on mass, energy, and digital information quantities of conservation-related topics produced or stored are presented in consecutive figures. Most of the technical characteristics of relative calculations are deployed in extensive captions of Figures and Tables.

However, to better situate their order of magnitude -and the time horizon-some preliminary data are necessary. First, the linear size of a bit of information is "larger" than of an average atom, i.e. 25×10^{-9} m vs~ 10^{-10} m, respectively. Second, the number of atoms on Earth is of the order of 10^{49}-10^{50}. Third, the mass of Earth is 5.972×10^{24} kg. And, fourth, according to Andrae and Edler (2015) scenarios, communication technologies could use as much as 51% of global electricity capacity by 2030. Two remarks are to be underlined here: (1) current total energy (power) needs on Earth are ca 18,5 TW; and, (2) 2030 is the time horizon for UN/SDGs (Sustainable Development Goals) to get implemented.

3.1. Estimation of digital mass-energy-information metrics for conservation-related topics

In this first part, Figures 1 and 2 present the evolution of the metrics N_{bits} and R_{bits} for the 9 conservation-related topics in 3 technological platforms of creation and storage of digital information, i.e., web sites/web pages, *Web of Knowledge/Science* (WoK/S) and e-books (kindle)/time-period. Figure 3 presents the metrics M_{info} and Q_{info}, and their relative R_{bits} for three topics, i.e., biodiversity, sustainability, and climate change during the period 1990–2020; notice that these topics are the flag-concepts of the meta-scientific revolution of the '80s that re-oriented environmental research and policy-making (Troumbis and Iosifidis 2020, Figure 2). Table 3 summarizes metrics on audio-visual information made available by popular platforms, i.e., *Netflix* and *You Tube* environmental channels.

The sample of indicative results presented in this section, despite their demonstrative sporadic and anecdotal character, confirms the technical feasibility of calculations on the mass-energy-information equivalence principle metrics in topics related to conservation, especially for digital data sources such as web sites, Web of Knowledge/Science and e-books (kindle).

At the same time, limitations and likely sources of data inaccuracy are uncovered, as initially assumed. Assumptions are necessary for various data sources and variables, e.g., the average number of web pages per web site. Further, standardized data for audio-visual platforms streaming are almost impenetrable. For instance, although general statistics on first-run original content released (hours/year) or the evolution of the number of worldwide subscribers of *Netflix* are published, they are limited in time depth. Indirect estimations might be searched after *Netflix* account of bandwidth use or global Internet traffic share (12.6% of total downstream volume for 2019) – which is significant since the total flow of data within the entire Internet is estimated by CISCO (2018) of the order of 174 EB/month for fixed Internet traffic and 41 EB/month for mobile Internet traffic, in the year 2020. Altogether, all themes videos accounted for 60.6% of total downstream volume worldwide in 2019.

Ratios [$\log_{10}R$] of the various metrics to respective total annual digital production on Earth offer valuable information. As expected, Rs are

Table 2. Summary of the structure of data sets used in the analysis and the calculations of various R_{bits}. See text for details and definitions.

Topics (digital production/storage)	#	Associated terms (extensions/specifications)	Technological components	Period/Years/Frequency
Total	All	All	All	1990-2020/30/year^{-1}
Conservation-related sub-sample				
Conservation-related topics	9			
1. Biodiversity	437			
2. Sustainability	527			
3. Biological conservation	502		Web sites/Web pages	1990-2020/30/year^{-1}
4. Climate change	542		Netflix (movies, docuseries, TV shows)	2020 (since 1999)/21 years
5. Pollution	524		Environmental Youtube channels (20)	2004 (first)/26/docs.day^{-1}
6. Ecology	514		Web of Knowledge/Science	1990-2020/30/5 years
7. Wildlife	542		e-books (kindle)	1990-2020/30/5 years
8. Endangered species	339			
9. Ecosystem services	257			
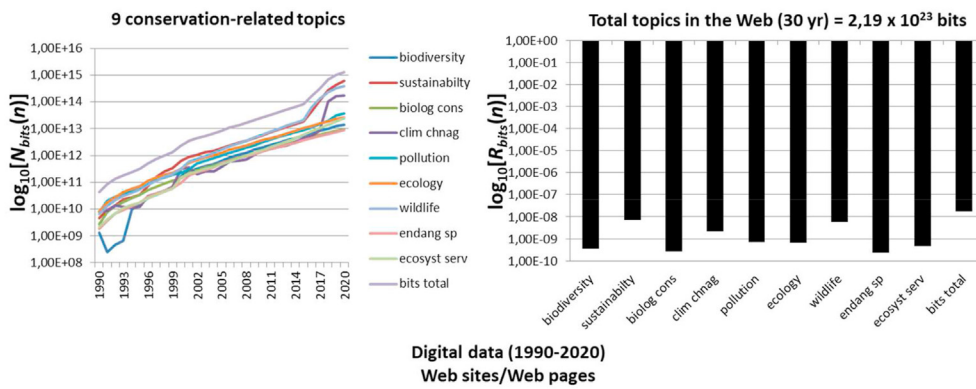
Figure 1. Indicative results (examples) of increase of digital data creation in conservation-related topics during time period 1990–2020 in web sites. Left panel: $\log_{10}$ scaled estimation of the total number of bits of information accumulated on the planet after n years of $f\%$ growth [$\log_{10}[N_{bits}(n)]$; the value of f growth rate used here is 10% (or 0.1). Right panel: $\log_{10}$ Ratios [$\log_{10} Rbits(n)$] of digital data creation per conservation-related topic during 1990–2020 divided by the total number of bits produced on Earth during the same period. Topics code: biodiversity; sustainability; biolog con: biological conservation; clim chang: climate change; pollution; ecology; wildlife; Endang sp: endangered species; ecosyst serv: ecosystem services. *Notice 1*: the term "bits total" in the panels corresponds to the sum of all studied conservation topics. *Notice 2*: Data for year 2020 refer to months January 1 – August 31. *Notice 3*: the average number of web pages/web site is fixed to 20 (on average: https://css-tricks.com/average-web-page-data-analyzing-8-million-websites/).
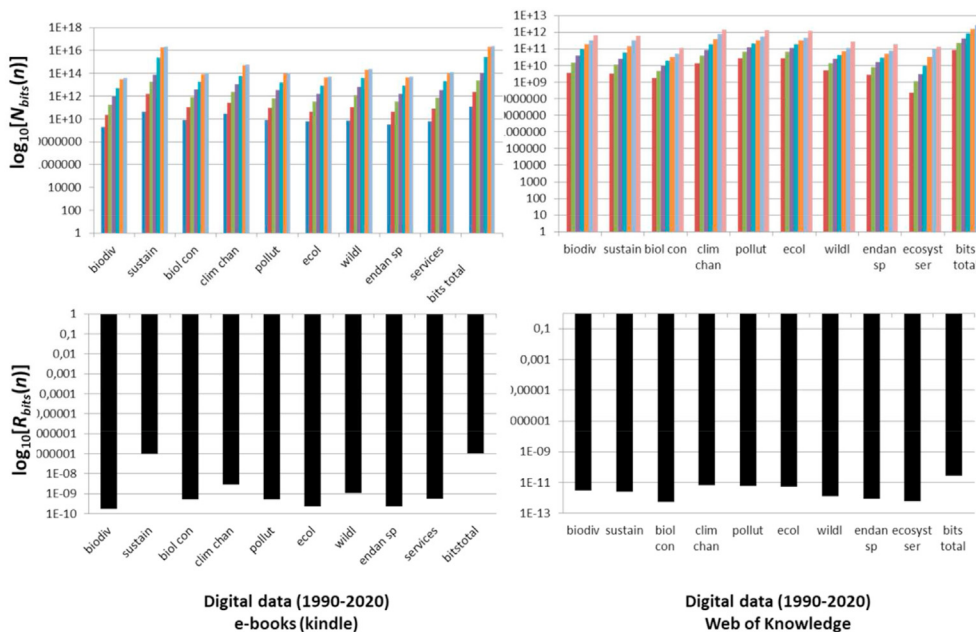
Figure 2. Indicative results (examples) of increase of digital data creation in conservation-related topics during time period 1990–2020 in e-books (kindle) and Web of Knowledge/Science. Study period is 1990–2020, presented here in 6 consecutive 5-year time classes, e.g. 1990–1995, 1996–2000, etc. Upper Row Panels: $\log_{10}$ scaled estimation of the total number of bits of information accumulated on the planet given $f\%$ growth [$\log_{10}[Nbits(n)]$; the value of f growth rate for digital data created in e-books (kindle) is 10% (or 0.1); the value of f growth rate for digital data created in Web of Knowledge/Science publications is 30% (or 0.3). Lower Row Panels: Indicative results (examples) of Ratios [$\log_{10} Rbits(n)$] of digital data creation per conservation-related topic during 1990–2020 divided by the total number of bits produced on Earth during the same period. Topics code: biodiv: biodiversity; sustain: sustainability; biol con: biological conservation; clim chan: climate change; pollut: pollution; ecol: ecology; wildl: wildlife; endan sp: endangered species; ecosyst serv: ecosystem services. Color code: blue: period 1990–95; red: 1996–00; green: 2001–05; purple: 2006–10; blue light: 2011–15; orange: 2016–20; grey: total of 30-year period. *Notice 1*: the term "bits total" in the panels corresponds to the sum of all studied conservation topics. *Notice 2*: Data for year 2020 refer to months January 1 – August 31.

<1 and their order of magnitude floors at ∼ 10^{-34} for audio-visual web streaming (*Netflix*, 2020) and ceils at ∼ 10⁻7 for e-books (kindle). Significant differences in *Rs* are observed between specific conservation-related topics: overall, sustainability heads the list of interest in text material followed by climate change, wildlife, and pollution. Metrics for all 9 conservation-related topics show an increase during the 1990–2020 time period and are positively influenced by the growth rate $f\%$.

3.2. Digital information metrics of net primary productivity and monetary value of Earth's biomes' ecosystem services

In the second part, calculations focus on energy-mass-information equivalence metrics and values of Net Primary Productivity (NPP) and the monetary value of Ecosystem Services (ES) of the various Earth's biomes. Data sources are presented in the captions of Figure 4.
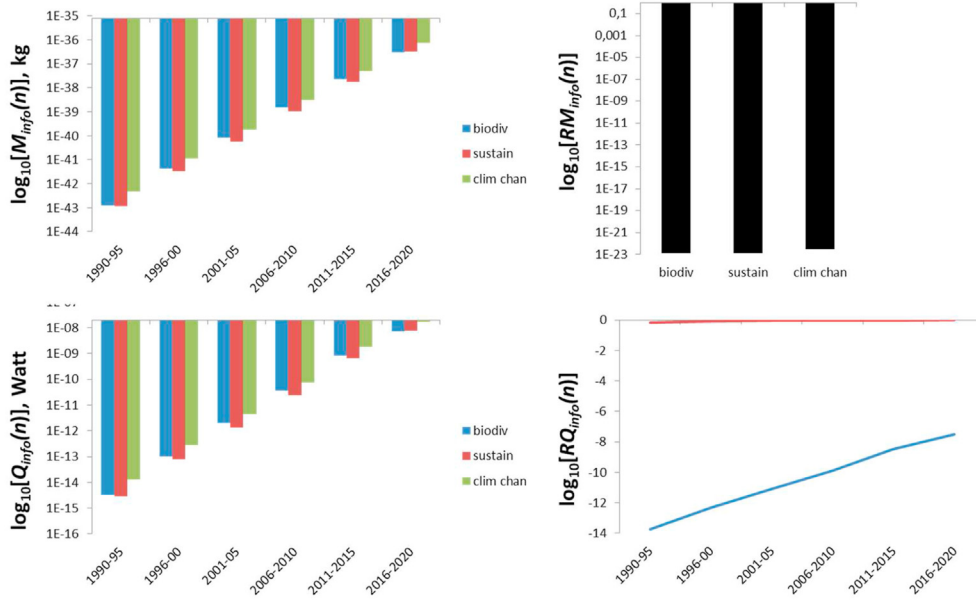
Figure 3. Indicative examples of results (and presentations) of accumulated mass of digital data and power required to generate and store the 3 emblematic conservation-related topics, i.e. biodiversity, sustainability and climate change, during time period 1990–2020, in the Web of Knowledge/Science. Left panels: $\log_{10}$ scaled estimation of digital mass of the 3 conservation-related information accumulated on the planet after 30 years with $f\% = 50\%$ [$\log_{10}[M_{info}(n)]$; $\log_{10}$ scaled estimation of power required to generate and store the 3 conservation-related information on the planet after 30 years with $f\% = 50\%$ [$\log_{10}[Q_{info}(n)]$]. Wright panels: $\log_{10}$ scaled Ratios [$\log_{10}[R_{info}(n)]$] for the same conservation-related topics and time period, respectively. *Notice 1*: Data for year 2020 refer to months January 1 – August 31. *Notice 2*: today's annual planetary power consumption is 18.5 TW. *Notice 3*: the average number of bytes and bits per Web of Knowledge/Science paper follows typical definitions in GreenNet: https://www.greennet.org.uk/support/understanding-file-sizes).

Table 3. Indicative results (examples) of Ratios [$\log_{10} R_{bits}(n)$] for audio-visual digital data creation in 9 conservation-related topics and time period divided by the annual rate of total digital bits production on Earth. *Notice 1*: *Netflix* streaming/movie starts in 1999 (first year of actual operation); Environmental docs refer to 20 best and more popular *You Tube* specialized channels. *Notice 2*: start of operation of various Environmental channels is different (e.g. *Earth Justice*, start: 2006 vs *Mongabay*, start: 2016) as different is the frequency of videos released/channel (e.g. *UN Environment*: 1 video/day vs. *The Conservation Fund*: 1 video/quarter).

| Audio-visual platforms	R_{bits} (total)	R_{bits} (2020)
Netflix (1999-): Movies, docuseries, TV shows	$2,88E^{-09}$	$4,90E^{-34}$
You Tube Environmental channels(aver: 8 years): Environmental documentaries	$3,15E^{-11}$	$3,34E^{-09}$

In this case, Ratios uncover indirectly the intensity and efficacy of research regarding major ecosystem services (example, NPP as the mass of C sequestrated in biomass) and their corresponding attributed monetary value on an annual basis. There are three comments of particular interest: (1) aquatic systems (both marine and terrestrial) are given a much higher value than terrestrial ones; especially, wetlands, marshes and swamps and, rivers and lakes, are comparable to the tropical forest; (2) the two series of data (NPP and ES) are not directly comparable because of the different classification schemes of biomes proposed in the literature; (3) intensity of research does not necessarily correspond to the efficacy of research, both indicated by Ratios involving M*info* as the denominator; on the contrary, several biomes with low NPP are given disproportionate high ES values in comparison to those heading the NPP order classification. This might mean that services other than supporting, e.g., provision, regulation, or cultural, are essential in human ES valuation hierarchies.

4. Discussion

Within the framework of the mass-energy-information equivalence principle, a bit of information is not an abstract mathematical entity; it is a physical entity, as stated by Vopson (2020). As a theory, the equivalence principle following Landauer's principle (1961) inoculates several physics domains (e.g., Tsallis 2019). In ecology and conservation sciences, the core concept of diversity of an ecological system is classically measured after Shannon-Wiener index based on information theory (e.g., Spellerberg and Fedor 2003).

In that sense, digital information is not just a privileged domain for high tech companies and startups; it is a valuable public commodity. Results, especially those relating to NPP and monetary value of ecosystem services, underline this property. However, observed orders of magnitude -especially for the estimated total mass of digital information- are 'insignificant' *per se*. However, time projections for accumulated digital mass and power requirements to affect the biosphere considerably show that such conditions are not away. It is predicted that - assuming the verisimilar $f\% \geq 50\%$ - energy requirements for digital production would consume half of the global electrical capacity in a few decades; and, in 1–2 centuries, digital mass accumulation would reach half of the Earth mass (Vopson 2020). Our analysis shows the lowest limits of the contribution of conservation-related digital information. Along with typical components of planetary change, i.e., extinction, land-use change, climate change, and degradation of physical and chemical quality of natural environment rates, time is not unlimited for holistic apprehension of new archetypes of conservation strategies, goals, and targets. Especially if Internet penetration rate and new users population increase drastically and ICT-based transitions towards a low carbon economy are not met with significant success, then the above challenges come closer in time.

Finally, a problem that has not been adequately revealed in the Methods and Results sections here above is that data provided by different digital bases and public domain services might diverge somehow. For example, results on the increase of the number of bits (estimation) for the 9 conservation-related topics (Figure 1) on digital data related to e-books (kindle) (Figure 2) retrieved from *Google Search* service
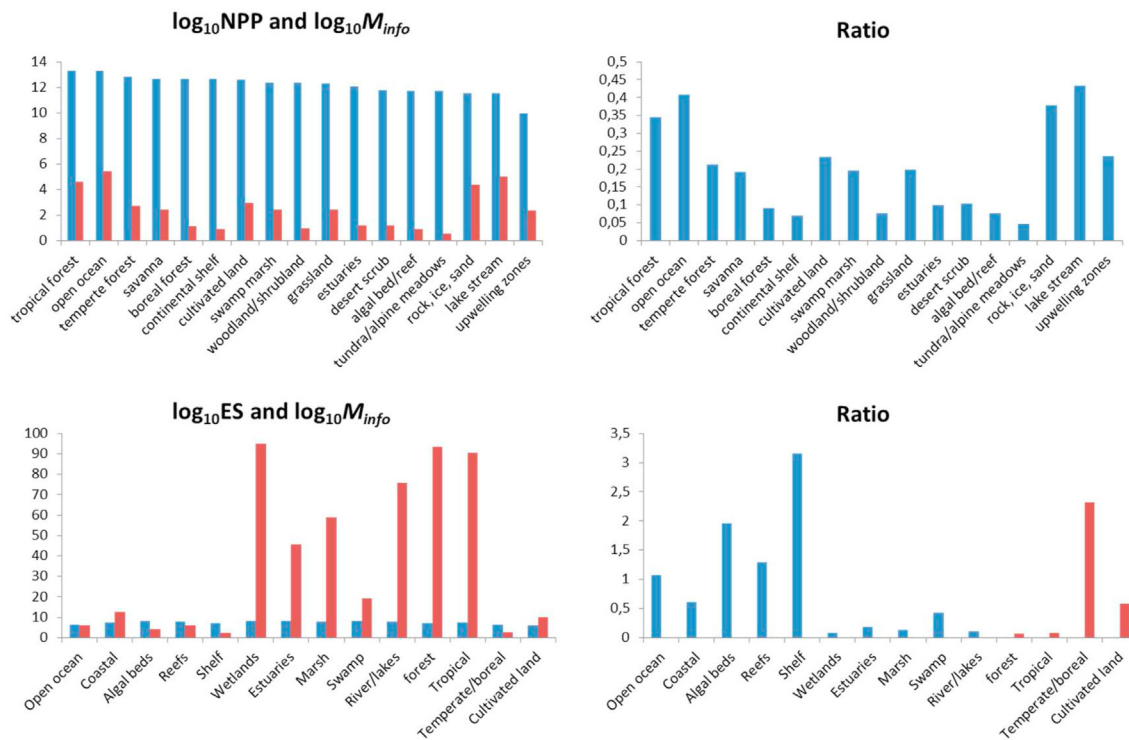
Figure 4. Indicative examples of digital information metrics of Net Primary Productivity and monetary value of Ecosystem Services of Earth's biomes. Upper panels: Net Primary Productivity (average) by surface area (kg(C)/m² /yr) and cover type (biomes), arranged in a decreasing value order; data of NPP are compiled after Vitousek et al. (1986) and Harte (1988). Color code: blue: $\log_{10}$NPP; red: $\log_{10}M_{info}$, accumulated digital information contained in Web of Knowledge/Science publications (1990–2020), with $f\%$ = 50%. Ratio: variation of ratios between NPP and accumulated digital information (in $\log_{10}$ values) across the range of Earth's biomes. Lower panels: Average total value of annual ecosystem services by biome type. Data provided and biomes are arranged according to Costanza's et al. scheme (1997). Color code: red: total value \$/area/year (area units used here: m² after transformation of hectares times 10^4); blue: $\log_{10}M_{info}$, accumulated digital information contained in Web of Knowledge/Science publications (1990–2020) in $\log_{10}$bits x 10^{-31}. Ratio: variation of ratios between annual value of ES and accumulated digital information contained in Web of Knowledge/Science publications (in $\log_{10}$ values) across the range of Earth's biomes. Color code: blue: aquatic (marine and terrestrial) ecosystems; red: terrestrial ecosystems.
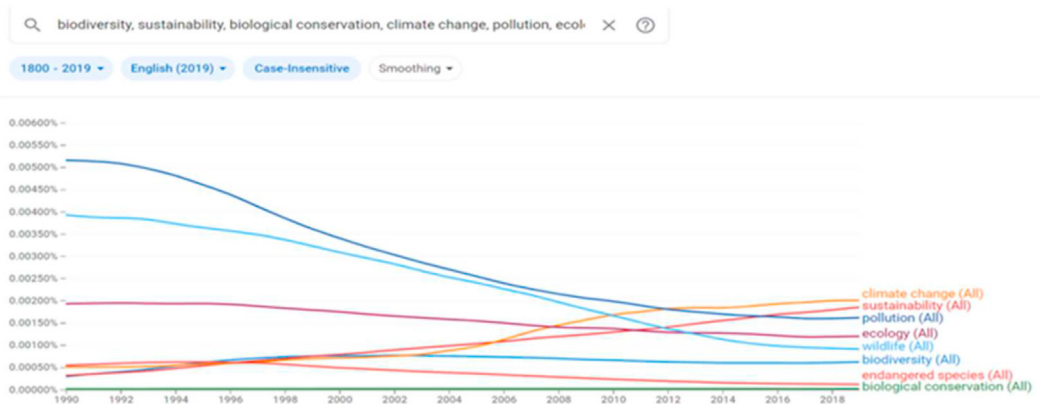
Figure 5. Extract from the dashboard of *Google Books Ngram Viewer*: frequencies of appearance (period 1990–2019) of the 9 selected conservation-related topics within 16 M digitized books.

are only partially matching data retrieved through *Google Books Ngram Viewer,* for the same period (Figure 5).

5. Conclusions

This perspective article attempts to present a novel, distinctive viewpoint on the quantification of conservation-related information grounded on the mass-energy-information equivalence principle and comparing limited evidence on digital data creation/time (9 conservation topics times 5 technological components, i.e., number of web sites/

web pages, publication in the *Web of Knowledge/Science*, e-books(kindle), audio-visual material in *Netflix* and *You Tube* environmental channels) to all digital data produced on Earth/same period of time. We assume that this thought-provoking line of argument represents an advance in implementing the conservation culturomics epistemology and its potential contribution to conservation problems and solutions. We examine the impact of conservation topics-related digital data by calculating their fraction of total digital data that humans produce. Deliberately, we used, as a precautionary assumption, the lowest values of available digital data statistics to avoid amplification of unnecessary noise in our findings.

However, there are three prerequisites, at least, to make this perspective genuinely operational. The first relates to the building of multidisciplinary and international academic networks, similar to those that gave birth to the conceptual revolution of the '80's on meta-scientific concepts, such as sustainability, biodiversity, or planetary/climate change (Bettencourt and Kaur 2011; Troumbis and Iosifidis 2020). Such academic communities should include conservation scientists, computer scientists, big data analysts and social media experts, linguists and lexicologists, etc. And, they have to build a certain prestige of this methodology to aspire for funding.

The second prerequisite refers to culturomics and conservation culturomics methods *per se*; they face inherent challenges, including developing procedures to enable dataset automation, acquisition, and management (Sherren et al., 2017; Roll et al., 2018). The thematic range of challenges is as complex as are the diverse interactions humans have with nature and its conservation (Sutherland et al., 2018). It extends roughly from differences in cultural perceptions of nature (Roll et al., 2016) to nature's effects on human well-being (IPBES Conceptual Framework: Diaz et al., 2015); from seasonal trends of human interest in natural phenomena (Mittermeier et al., 2019) to longer trends in public interest in conservation issues in general (Proulx et al., 2013; Burivalova et al., 2018; Legagneux et al., 2018; Correia et al., 2019); or, from mismatches between scientific effort(s) and conservation needs (Fisher et al., 2010) to public valuation(s) of biodiversity and landscapes (Roberge 2014; Correia et al. 2016, 2017; Davies et al., 2018).

The third prerequisite refers to the inclusion of conservation-related digital data created in the social media sphere, e.g., *Facebook, Twitter, Flickr*, etc. It is not only a necessary addition of this big source of digital data; it is methodologically required to use them in sentiment analysis, a fundamental element in dealing with the misrepresented controversies, especially the "new information controversy". Sentiment analysis has been widely applied for both scientific and commercial applications. For instance, the VADER sentiment analysis technique (e.g., Urologin 2018) has been used to study and promote conservation (Lennox et al., 2019; Toivonen et al., 2019).

Further, big corporations and Governments are the gatekeepers of powerful datasets; they are, almost systematically, reluctant to publish digital data statistics beyond their "legal accounting obligations" (e.g. sales, market share or growth in users' numbers): a reader might gain a better understanding of the process watching the US Congressional antitrust hearing of CEOs of *Amazon, Google, Apple,* and *Facebook*, July 30, 2020 (You tube: https://www.youtube.com/watch? v = XzqrwsPtLxE). For instance, *Amazon* does not share its numbers and hence there are not accurate industry statistics on e-books (kindle). There are just guesses on 1 million e-books newly published every year, and that is in English language only and mostly in the US market (https://www.quora.com/How-many-eBooks-are-published-every-year-Are-they-exceeding-print-books-thanks-to-self-publishing-tools). In the same line, *Google* does not disclose the minimum threshold of searches used to normalize *Google Trends* rank percentages (Burivalova et al., 2018). This condition imposes significant efforts towards appropriate and equitable use of these tools while understanding their inherent biases concerning societal sections they represent (Funk and Rusowsky 2014; Ladle et al., 2016; Troumbis and Iosifidis 2020).

Finally, technical advances in the field of conservation culturomics are to be imminently expected in developing more elaborate techniques of treatment of data series on public interest, awareness, and engagement in conservation-related terms. However, the core challenge is epistemic per se, i.e., shifting from a posteriori analysis of data to testable a priori declarations and hypotheses. The use of analysis of social networks information diffusion and 'rumor'-like spreading models might be a promising way for the construction of a true scientific paradigm in culturomics.

Declarations

References

Aiden, E., Michel, J.-B., 2013. Uncharted: Big Data as a Lens on Human Culture. Riverhead Books.

Andrae, A., Edler, T., 2015. On global electricity usage of communication technology: trends to 2030. Challenges 6 (1), 117–157.

Bettencourt, L.M.A., Kaur, J., 2011. Evolution and structure of sustainability science. PNAS 108 (49), 19540–19545.

Blicharska, M., Orlikowska, E.H., Roberge, J.-M., Grodzinska-Jurczak, M., 2016. Contribution of social science. to large scale biodiversity conservation: a review of research about the Natura 2000 network. Biol. Conserv. 199, 110–122.

Bridge, G., Bouzarovski, S., Bradshaw, M., Eyre, N., 2013. Geographies of energy transition: space, place and the low-carbon economy. Energy Pol. 53, 331–340.

Burivalova, Z., Butler, R.A., Wilcove, D.S., 2018. Analyzing Google search data to debunk myths about the public's interest in conservation. Front. Ecol. Environ. 16 (9), 509–514.

CISCO, 2016. Global - 2020 Forecast Highlights. https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2020_Forecast_Highlights.pdf.

CISCO, 2018. Cisco Visual Networking Index: Forecast and Trends, 2017–2022. (Accessed 22 October 2020).

Clausen, R., York, R., 2008. Global biodiversity decline of marine and freshwater fish: a cross-national analysis of economic, demographic, and ecological influences. Soc. Sci. Res. 37 (4), 1310–1320.

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. Biol. Conserv. 230, 29–36.

Correia, R.A., Jepson, P.R., Malhado, A.C.M., Ladle, R.J., 2016. Familiarity breeds content: assessing bird species popularity with culturomics. PeerJ 4, e1728.

Correia, R.A., Jepson, P., Malhado, A.C.M., Ladle, R.J., 2017. Internet scientific name frequency as an indicator of cultural salience of biodiversity. Ecol. Indicat. 78, 549–555.

Correia, R.A., Di Minin, E., Jarić, I., Jepson, P., Ladle, R., Mittermeier, J., Roll, U., Soriano-Redondo, A., Veríssimo, D., 2019. Inferring public interest from search engine data requires caution. Front. Ecol. Environ. 17, 254–255.

Costanza, R., d'Arge, R., de Groot, R., Farberk, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260.

Dahl, R., 2010. Greenwashing: do You know what you're buying? Environ. Health Perspect. 118 (6), a246–a252.

Davies, T., et al., 2018. Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment. PloS One 13, e0203694.

Diamond, J., 2011. Collapse. How Societies Choose to Fail or Survive, second ed. Penguin Random House, UK.

Diaz, S., Demissew, S., Carabias, J., Joly, C., et al., 2015. The IPBES Conceptual Framework - connecting nature and people. Curr. Opin. Environ. Sustain. 14 (1), 1–16.

Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front. Environ. Sci. 3, 63.

Fisher, R., Radford, B.T., Knowlton, N., Brainard, R.E., Michaelis, F.B., Caley, M.J., 2010. Global mismatch between research effort and conservation needs of tropical coral reefs. Cons. Lett. 4, 64–72.

Floridi, L., 2014. The Fourth Revolution: How the Infosphere Is Shaping Human Reality. Oxford University Press, Oxford.

Funk, S.M., Rusowsky, D., 2014. The importance of cultural knowledge and scale for analysing internet search data as a proxy for public interest toward the environment. Biodivers. Conserv. 23, 3101–3112.

GeSI –Global e-Sustainability Initiative, 2008. SMART 2020: Enabling the Low Carbon Economy in the Information Age. Technical Report, p. 85.

Harte, J., 1988. Consider a Spherical Cow. A Course in Environmental Problem Solving. University Science Books, Sansalito, California.

Kim, J.Y., Do, Y., Im, R.-Y., et al., 2014. Use of large web-based data to identify public interest and trends related to endangered species. Biodivers. Conserv. 23 (12), 22961–22984.

Ladle, R., Correia, R.A., Do, Y., Joo, G.-J., Malhado, A., et al., 2016. Conservation culturomics. FEE 14 (5), 269–275.

Ladle, R.J., Jepson, P., Correia, R.A., Malhado, A.C.M., 2019. A culturomics approach to quantifying the salience of species on the global internet. People Nat.

Landauer, R., 1961. Irreversibility and heat generation in the computing process. IBM J. Res. Dev. 5 (3), 183–191.

Landauer, R., 1996. The physical nature of information. Phys. Lett. A 217 (4-5), 188–193.

Laufer, W.S., 2003. Social accountability and corporate greenwashing. J. Bus. Ethics 43 (3), 253–261.

Legagneux, P., et al., 2018. Our house is burning: discrepancy in climate change vs. Biodiversity coverage in the media as compared to scientific literature. Front. Ecol. Evol. 5, 175.

Lennox, R.J., Veríssimo, D., Twardek, W.M., Davis, C.R., Jarić, I., 2019. Sentiment analysis as a measure of conservation culture in scientific literature. Conserv. Biol.

McCallum, M.L., Bury, G.W., 2013. Google search patterns suggest declining interest in the environment. Biodivers. Conserv. 22, 1355–1367.

McCallum, M.L., Bury, G.W., 2014. Public interest in the environment is falling: a response to Ficetola (2013). Biodivers. Conserv. 23 (4), 1057–1062.

McKinsey, Company, 2009. Pathways to a Low Carbon Economy: Global Greenhouse Gas Abatement Cost Curve. Technical Report, p. 189.

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 Biol. 17, e3000146.

Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Aiden, E., 2011. Quantitative analysis of culture using millions of digitized books. Science 331, 176–182.

Nghiem, L.T.P., Papworth, S.K., Lim, F.K.S., Carrasco, L.R., 2016. Analysis of the capacity of Google Trends to measure interest in conservation topics and the role of online news. PloS One.

Pascual, U., Balvanera, P., Díaz, S., Pataki, G., et al., 2017. Valuing nature's contributions to people: the IPBES approach. Curr. Opin. Environ. Sustain. 26–27, 7–16.

Proulx, R., Massicotte, M., Pepino, M., 2013. Googling trends in conservation biology. Conserv. Biol. 28, 44–51.

Ramus, C.A., Montiel, I., 2016. When are corporate environmental policies a form of greenwashing? Bus. Soc. 44 (4), 377–414.

Roberge, J.-M., 2014. Using data from online social networks in conservation science: which species engage people the most on Twitter? Biodivers. Conserv. 23, 715–726.

Robinson, A., Schlamminger, S., 2016. The watt or Kibble balance: a technique for implementing the new SI definition of the unit of mass. Metrologia 53 (5), A46–A74.

Roll, U., Mittermeier, J.C., Diaz, G.I., Novosolov, M., Feldman, A., Itescu, Y., Meiri, S., Grenyer, R., 2016. Using Wikipedia page views to explore the cultural importance of global reptiles. Biol. Conserv. 204 (Part A), 42–50.

Roll, U., Correia, R.A., Berger-Tal, O., 2018. Using machine learning to disentangle homonyms in large text corpora. Conserv. Biol. 32, 716–724.

Sherren, K., Smit, M., Holmlund, M., et al., 2017. Conservation culturomics should include images and a wider range of scholars. FEE 15 (6), 289–290.

Spellerberg, I.F., Fedor, P.J., 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Global Ecol. Biogeogr. 12 (3), 177–179.

Sutherland, W.J., Butchart, S.H.M., Connor, B., et al., 2018. A 2018 Horizon Scan of Emerging Issues for Global Conservation and Biological Diversity. TREE.

Terborgh, J., 1999. Requiem for Nature. Island Press.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: a methodological overview. Biol. Conserv. 233, 298–315.

Troumbis, A.Y., 2017a. Declining Google Trends of public interest in biodiversity: semantics, statistics or traceability of changing priorities? Biodivers. Conserv. 26 (6), 1495–1505.

Troumbis, A.Y., 2017b. Google Trends and cycles of public interest in biodiversity: the animal spirits effect. Biodivers. Conserv. 26 (14), 3421–3443.

Troumbis, A.Y., 2019. The time and timing components of conservation culturomics cycles and scenarios of public interest in the Google era. Biodivers. Conserv. 28 (7), 1717–1727.

Troumbis, A.Y., Hatziantoniou, M.N., 2018. Too much, too fast, too complex or too strange? Asymmetric sequences in public opinion regarding biodiversity conservation in Island social-ecological setups. Biodivers. Conserv. 27 (6), 1403–1418.

Troumbis, A.Y., Iosifidis, S., 2020. A decade of Google Trends-based Conservation culturomics research: a critical evaluation of an evolving epistemology. Biol. Conserv. 248, 108647.

Tsallis, C., 2019. Beyond Boltzmann-Gibbs-Shannon in physics and elsewhere. Entropy 21 (7). Article Number: 696.

Urologin, S., 2018. Sentiment analysis, visualization and classification of summarized news articles: a novel approach. Int. J. Adv. Comput. Sci. Appl. 9 (8), 616–625.

Vitousek, P.M., Ehrlich, P.R., Ehrlich, A.H., Matson, P.A., 1986. Human appropriation of the products of photosynthesis. Bioscience 36 (6), 368–373.

Vopson, M.M., 2019. The mass-energy-information equivalence principle. AIP Adv. 9, 095206.

Vopson, M.M., 2020. The information catastrophe. AIP Adv. 10, 085014.

Wehn, U., Almomani, A., 2019. Incentives and barriers for participation in community-based environmental monitoring and information systems: a critical analysis and integration of the literature. Eur. Sci. Pol. 101, 341–357.

Zikopoulos, P., deRoos, D., Parasuraman, K., Deutsch, T., Giles, J., Corrigan, D., 2012. Harness the Power of Big Data: the IBM Big Data Platform. McGraw-Hill Professional, New York.