Trends in Ecology & Evolution

CellPress
REVIEWS

Review

# iEcology: Harnessing Large Online Resources to Generate Ecological Insights

Ivan Jarić,[1,2,*] Ricardo A. Correia,[3,4,5,6] Barry W. Brook,[7,8] Jessie C. Buettel,[7,8] Franck Courchamp,[9] Enrico Di Minin,[3,4,10] Josh A. Firth,[11,12] Kevin J. Gaston,[13] Paul Jepson,[14] Gregor Kalinkat,[15] Richard Ladle,[6] Andrea Soriano-Redondo,[16,17] Allan T. Souza,[1] and Uri Roll[18]

**Digital data are accumulating at unprecedented rates. These contain a lot of information about the natural world, some of which can be used to answer key ecological questions. Here, we introduce iEcology (i.e., internet ecology), an emerging research approach that uses diverse online data sources and methods to generate insights about species distribution over space and time, interactions and dynamics of organisms and their environment, and anthropogenic impacts. We review iEcology data sources and methods, and provide examples of potential research applications. We also outline approaches to reduce potential biases and improve reliability and applicability. As technologies and expertise improve, and costs diminish, iEcology will become an increasingly important means to gain novel insights into the natural world.**

## Highlights

iEcology is a new research approach that seeks to quantify patterns and processes in the natural world using data accumulated in digital sources collected for other purposes.

iEcology studies have provided new insights into species occurrences, traits, phenology, functional roles, behavior, and abiotic environmental features.

iEcology is expanding, and will be able to provide valuable support for ongoing research efforts, as comparatively low-cost research based on freely available data.

We expect that iEcology will experience rapid development over coming years and become one of the major research approaches in ecology, enhanced by emerging technologies such as automated content analysis, apps, internet of things, ecoacoustics, web scraping, and open source hardware.

## Information Age, Big Data, and iEcology

The information age is characterized by rapid accumulation of myriad types of digital data [1]. Central to this revolution is the Internet, which is a source of unprecedented amounts of diverse and readily accessible data, via webpages, social media, and various other data platforms. These data are constantly created and stored in the digital realm and form an omnipresent part of the modern world. They also provide novel opportunities for research that the scientific community is only beginning to explore. Here, we describe an emerging research approach – iEcology (i.e., internet ecology), which we define as the study of ecological patterns and processes using online data generated for other purposes and stored digitally (Figure 1). These data can be used to address fundamental ecological questions and to analyze ecological processes at a range of spatiotemporal scales and across a diverse range of contexts. As such, iEcology has the potential to provide new understandings of ecological dynamics and mechanisms, complementing more traditional methods of obtaining ecological data.

While iEcology can be considered to fit within the wider scope of **ecological informatics** (see Glossary), it is distinct from other uses of **Big Data** sources in the biological sciences in that data are not specifically and intentionally generated to address ecological and environmental questions [2–4]. Moreover, iEcology expands on the traditional scope of ecological informatics with new data sources and dedicated methods to analyze them. iEcology is predominantly focused on collecting, collating, and exploring data generated online by human society, either passively or unintentionally (e.g., Internet search activity, social media interactions, and uploaded data and media), a process also referred to as passive crowdsourcing [5]. iEcology uses digital methods to access, handle, and analyze these data, in a manner akin to techniques from other research fields such as sociology, culture and media studies, biomedical sciences, computer sciences, and economics [6,7]. iEcology also shares part of its toolbox with **conservation culturomics** – an emerging research area in conservation science [8–10] – albeit with a different focus. Specifically, while conservation culturomics is

[1]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic
[2]University of South Bohemia, Faculty of Science, Department of Ecosystem Biology, České Budějovice, Czech Republic
[3]Helsinki Laboratory of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland
[4]Helsinki Institute of Sustainability Science, University of Helsinki, Helsinki, Finland
[5]Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal
[6]Institute of Biological and Health Sciences, Federal University of Alagoas, Av. Lourival Melo Mota, Maceió, AL, Brazil

https://doi.org/10.1016/j.tree.2020.03.003

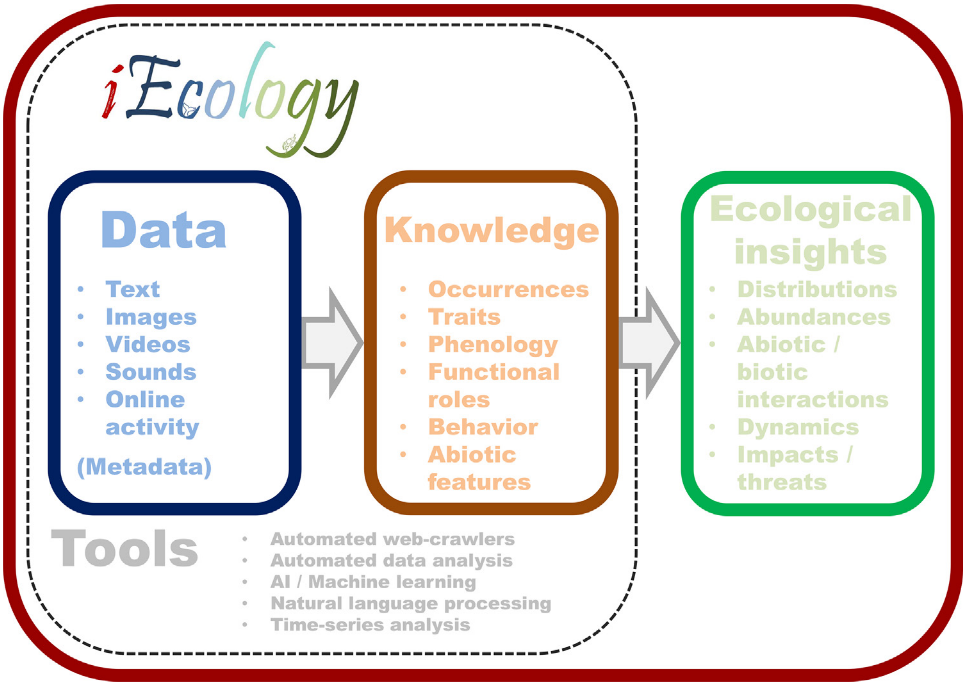


Figure 1. Conceptual Representation of iEcology, Showing How Key Data Types Can Turn into Knowledge of the Natural World Using a Set of Research Tools, Which in Turn Can Provide Novel Ecological Insights. Abbreviation: AI, artificial intelligence.

[7]School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia
[8]ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Tasmania, Hobart, Tasmania, Australia
[9]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, 91405, Orsay, France
[10]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa
[11]Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, UK
[12]Merton College, University of Oxford, Oxford, UK
[13]Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall, UK
[14]School of Geography and the Environment, University of Oxford, Oxford, UK
[15]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
[16]Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Universidade do Porto, Vairão, Portugal
[17]Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon, Portugal
[18]Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

*Correspondence:
ivan.jaric@hbu.cas.cz (I. Jarić).

interested in understanding human engagement with nature, iEcology methods focus on the ecological knowledge that can be gained from these human–nature interactions in the digital realm. iEcology data predominantly give rise to insights that are correlative in nature, similar to other large-scale ecological explorations such as much of macroecology [11], and should be viewed as such.

Here, we present a broad overview and description of iEcology, including its scope, data types, sources and methods, as well as current major caveats and future prospects for the development of this emerging research approach.

## Research Scope

Several recent studies have highlighted the potential of iEcology (Figure 2). The most common applications of such methods have been to explore species occurrences and their spatiotemporal trends (Figure 3). For example, a study comparing real-world encounter rates of bird species in the USA with Google Trends data found good agreement between the two sources (Figure 2A) [12]. This showcases the potential of using voluminous search engine data to explore species distributions in many regions. Others have explored species occurrences and distributions using various sources, such as Flickr, news articles, Twitter, YouTube, Facebook, and Google Trends [13–25], as well as population dynamics and phenology [14,20,23,26–31]. A particular illustration comes from assessing seasonal migration patterns of sockeye salmon (*Oncorhynchus nerka*) and Atlantic salmon (*Salmo salar*) from Wikipedia pageview frequencies (Figure 2B) [32]. In addition to mapping the distribution and occurrences of known species, images uploaded on social media have also been used to identify new species [33,34]. Trait dynamics, evolutionary trends, and biogeographic patterns can also be explored using iEcology methods. For instance, Google

Images were used to identify the presence and distribution of hybrid zones of hooded (*Corvus cornix*) and carrion (*Corvus corone*) crows in Europe (Figure 2C) [35]. Furthermore, spatiotemporal dynamics of biophysical environments, such as solar radiation and various other climatic parameters were characterized using Flickr tags [18].

iEcology sources, tools, and methods can also be used to explore biotic and abiotic interactions within and across species and their environments. For example, feeding patterns of yellow anaconda (*Eunectes notaeus*) and green anaconda (*Eunectes murinus*) were studied using online videos [36], while online images that simultaneously depicted African birds and herbivorous mammals were used to construct a web of associations between these two groups (Figure 2D) [37]. iEcology also provides new opportunities to study animal behavior [15]. For instance, YouTube videos have been used to compare the behavior of red (*Sciurus vulgaris*) and grey squirrels (*Sciurus carolinensis*) in different habitats (Figure 2E) [38]. The sheer volume and coverage of such sources could also prove fertile ground for identifying and tracking the spread of new behaviors [39–41]. Disease ecology, including knowledge of the occurrence, distribution, prevalence, and severity of diseases, has also recently benefited from iEcology methods [42].

iEcology methods have also been used to investigate ecosystem and habitat dynamics in response to increasing anthropogenic impacts. For example, videos of the Tour of Flanders cycling race from over 35 years have been used to track phenological changes to vegetation in response to climate change (Figure 2F) [27]. Images of corals and tweets referring to corals have both been used to evaluate the state and trends of coral reefs in different areas, suffering from various human impacts [43,44]. Aspects of invasion dynamics [14,45] and overexploitation of fish [29,30] have also been studied using image analysis, tweets, and news articles. In the same way, behavioral changes in animals in response to anthropogenic impacts [46–48] can be tracked by such methods.

While inherently varied in scope, other fields within ecology and environmental science could conceivably benefit from iEcology tools and methods, such as functional ecology, macroecology, landscape ecology, and urban ecology.

## iEcology Research Toolbox – Data Types, Sources, and Methods

At their core, iEcology data sources fall into two categories: (i) new data uploaded by users for different purposes; and (ii) data on online activity, including data access and search engine usage. Types of data within the first category can comprise text, images, videos, and sounds (Figure 1). The second category is aggregated data and the exploration of frequencies (e.g., the number of times a term was searched or a webpage visited, but could also include interactions on social media such as shares and likes). Both categories have different types of associated **metadata** that are particularly important for iEcology, such as locality, **timestamp**, user identity, and links across data.

iEcology data sources differ in their scope, availability, ease of access, associated metadata, and therefore utility for different types of research. Potential data sources range from various social media platforms (e.g., Twitter and Flickr) [49], search engines (e.g., Google, Baidu, and Bing), online encyclopedias (e.g., Wikipedia and Encyclopedia Britannica online), and other online repositories (blogs, discussion forums, popular articles, books, etc.). Many of these sources can also be accessed through search engines. The scope of sources differs based on spatiotemporal coverage, linguistic or cultural breadth, data resolution, and the degree of multimedia composition (e.g., text, images, and video) per source. Data also differ in availability: while many sources are freely available, some platforms may restrict availability by limiting data collection (i.e., limits on

## Glossary

**Application programming interface (API):** a *set* of protocols and tools that provide a communication interface between applications, commonly between a client and the server, to enable simplified construction of client-side software.

**Augmented reality:** digital devices that electronically supplement elements of the physical world with portable, interactive, computer-generated attributes, and synthetic sensory inputs.

**Big Data:** large volumes (petabyte scale upwards) of structured or unstructured data that require advanced tools for management and processing. Also denotes a unique field that explores methods to store, analyze, and generally deal with voluminous data that is too complex to be handled with traditional processing methods.

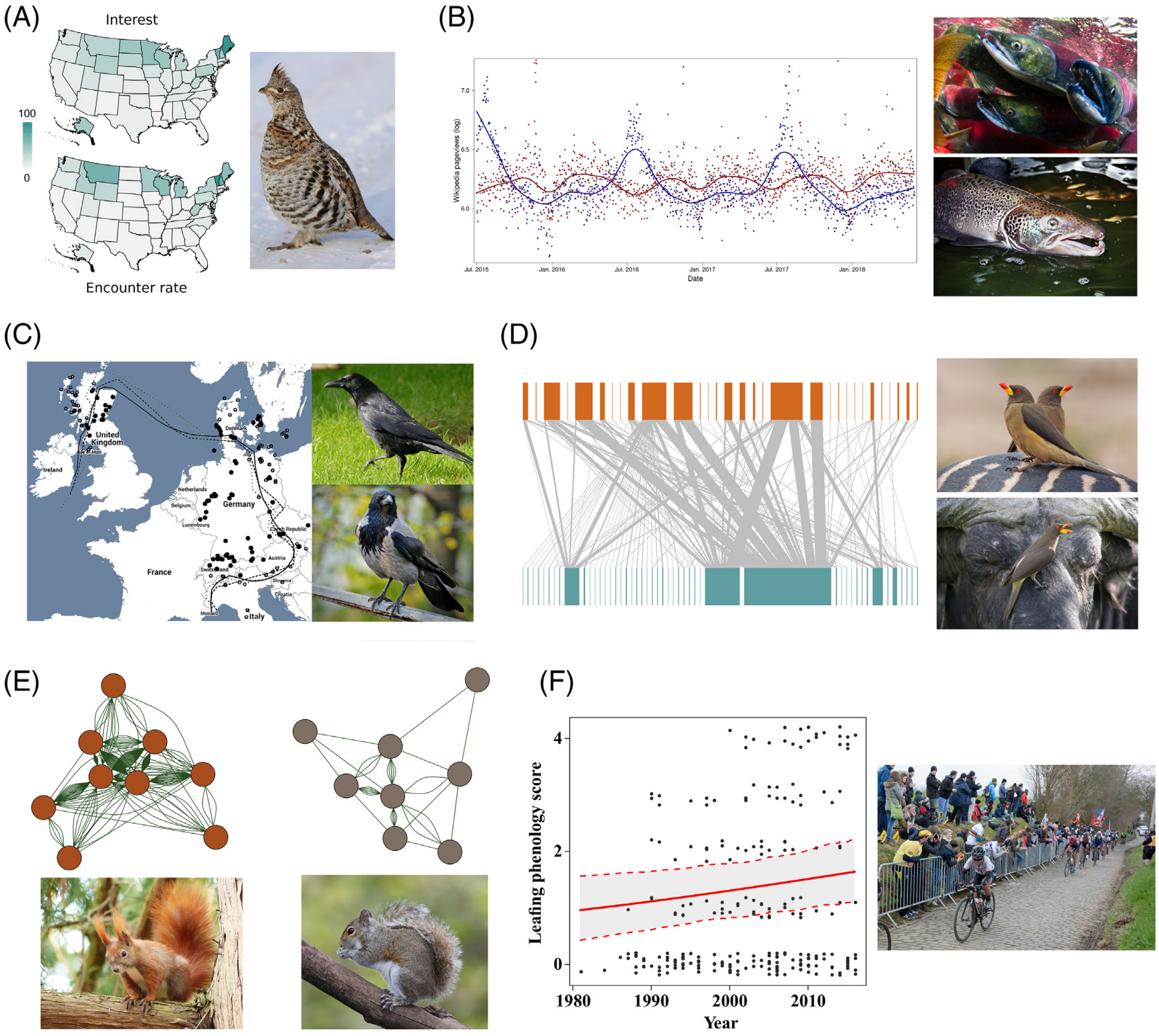
**Conservation culturomics:** emerging area of study that explores human interaction with nature through the quantitative analysis of digital data.

**Ecological informatics:** interdisciplinary framework for the management, analysis, and synthesis of ecological data by advanced computational technology [81].

**Ground truthing:** a process of obtaining information by direct observation, used as empirical evidence to test or validate inferred information.

**Metadata:** a set of data that provides basic information about other data, and facilitates tracking and processing. Examples include timestamps, geotags, information on data providers, data type, or precision.

**Timestamp:** a digital record of time that defines when an event has occurred.



Trends in Ecology & Evolution

Figure 2. Examples of iEcology Studies. (A) High level of correlation observed for ruffed grouse (*Bonasa umbellus*) encounter rate and spatial distribution of societal interest, based on Google Trends [12]. (B) Sockeye salmon (*Oncorhynchus nerka*, blue line; upper photo) and Atlantic salmon (*Salmo salar*, red line; lower photo) popularity based on Wikipedia pageviews reflects their seasonal migration patterns [32]. (C) Distribution of two crow species in Europe, carrion crow (*Corvus corone*; upper photo) and hooded crow (*Corvus cornix*; lower photo), indicated by Google Images, corresponds well with their actual distribution and hybrid zones [35]. (D) Quantitative bird–mammal association webs for non-oxpecker and oxpecker species in African birds and herbivorous mammals revealed by the analysis of Google Images [37]; upper photo – yellow-billed oxpeckers (*Buphagus africanus*) on zebra (*Equus* sp.), lower photo – yellow-billed oxpecker on Cape buffalo (*Syncerus caffer*). (E) The network visualization of the behavior of red (*Sciurus vulgaris*; left) and eastern grey (*Sciurus carolinensis*; right) squirrels assessed by YouTube videos [38]. (F) Phenological changes in vegetation as a response to climate change identified through archive videos of the Tour of Flanders cycling race [27]. See the Supplemental Information online for image attributions.

volume, time frame, or number of queries) or use (e.g., privileged access or paywall restrictions). Sources also differ in their ease of access, from simple online tools embedded at the source (e.g., Google Trends webpage), through open **application programming interfaces (APIs)** accessible via various dedicated computer scripts (e.g., Wikipedia and Flickr), to APIs with

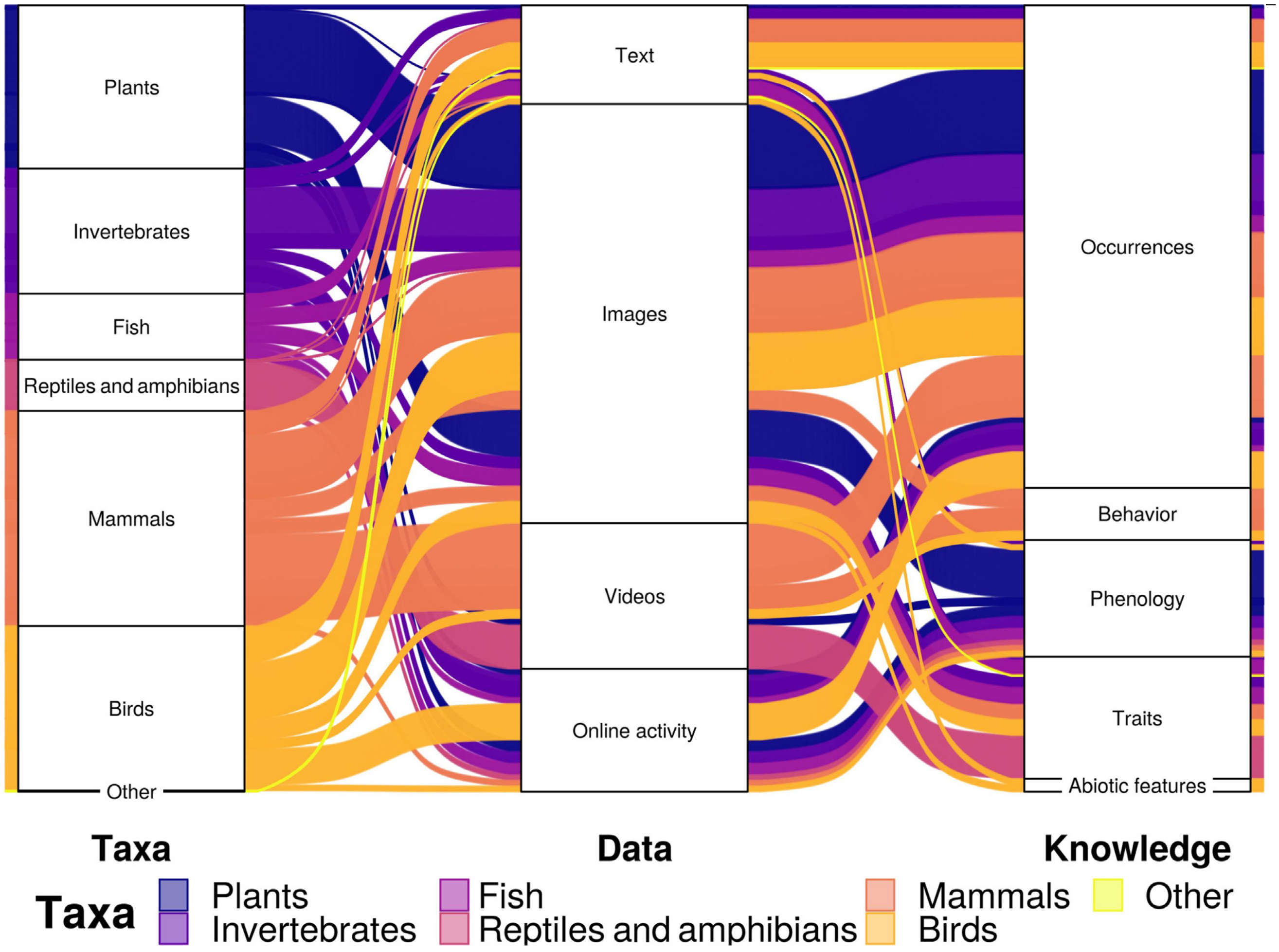


Figure 3. Overview of the Studied Taxa, Data Sources Used, and Knowledge Categories Addressed by the iEcology Studies Cited in this Article. Colors represent different taxa, width of lines represents relative number of publications connecting different categories.

restricted access (e.g., Facebook). However, data availability and ease of access to different sources can also change over time.

The analysis of iEcology data faces similar challenges and uses the same solutions as many other approaches for analysis of Big Data [2,50]. Many of the methods used in iEcology rely on high levels of automation, frequently adopting machine-learning techniques [51]. There are different tools that can aid each stage of the research: data access, downloading, handling, extraction, storage, pattern identification and recognition, data analysis, and visualization. These tools are in a constant state of evolution, as illustrated by developments in deep neural network analysis and other emerging technologies (Box 1).

## Caveats and Solutions

While holding remarkable promise, iEcology is subject to several inherent challenges and gaps that require careful consideration when undertaking such research (see Outstanding Questions). Primarily, it is important always to keep in mind that while ever increasing, the digital realm only encompasses a

### Box 1. Emerging Technologies Relevant for iEcology

Future development within iEcology will be enhanced by rapidly developing technologies:

Apps and Games

Apps on mobile devices are ubiquitous, and often within a person's reach 24-7. Using these to support **augmented reality** could also provide an interface to generate more detailed data with real-time diagnostics [64–66]. In addition, apps that 'gamify' nature can motivate the public to interact with their environment, and thus provide more data on species and the environment. Overall, apps and games have the potential to transform how humans interact with nature (both positive and negative) and cause a fundamental shift in the quantity and quality of iEcology data.

Automated Content Analysis

The application of algorithms for analyzing visual, textual, and audio content from digital sources. These methods have allowed, for example, automatic identification, counting, and description of species and individuals from images and videos [67,68], and the extraction from text of information on species and their interactions [69]. Further developments will allow combining visual, textual, and audio analysis of large volumes of iEcology data [70]. All these methods should be used carefully and consider ethical concerns [71].

Bioacoustics and Ecoacoustics

The recording and analysis of sounds produced by biological entities and entire environments. Increase in sonic and video recording and publicized soundscapes could provide an untapped source of data for iEcology [72–75].

Blockchain

Cryptographically linked and growing data lists. Further development of dedicated iEcology blockchains or plug-ins will allow the creation of immutable complex data of various formats that will be permanently recorded into a decentralized platform at the moment of their creation. This would increase security, traceability, decrease errors associated with multiple data entries, and allow imprinting of the technical details of data generator [76].

Internet of Things

A network of computers, machines, and other objects that share information and interact. This will greatly increase the amount of data pertaining to humans and their actions [77].

Open Source Hardware

Physical objects with design specifications that allow them to be widely studied, modified, created, and distributed. As more knowledge and expertise on construction of various sensors are produced and shared, larger volumes of high-quality and more specialized data could be produced [78,79].

Web Scraping

The fetching and extraction of relevant information from web content, mostly done automatically. Further developments in these technologies will enable better and quicker access to larger volumes of iEcology data, and potentially continuous monitoring of patterns and trends [80].

subset of the world – one that is nonrandom in extent and depth. Indeed, as the data are not generated systematically, there is great variance in content generation among different users, regions, cultures, and time frames, with inherent risks of biases [52]. Such individual and cultural subjectivity can further complicate data interpretation. Moreover, multiple entries of the same data by single or different users could cause biases related to nonindependence. Therefore, underlying data for iEcology research should neither be treated as randomly distributed, nor used in raw form without addressing these issues. Indeed, rather than ignoring such considerations, specific investigations into aspects of the data such as the nonrandom distribution and the level of nonindependence can actually provide further insights into data structure and any discovered patterns.

Several approaches, many already recognized within other fields of research that rely on online data, can be used to tackle these challenges. Validation with common and reputable sources such as systematic surveys, remote sensing, and citizen science (i.e., **ground truthing**) can decrease the level of associated uncertainty and help reinforce confidence in the data and their interpretation [12,44,53,54]. This is particularly important when testing new tools or approaches.

The vast majority of iEcology studies that we have identified have used multiple data sources to validate results, including data from field research, citizen science, online databases, scientific literature, or their combination [12,13,16–25,31,32,35,43,44]. In most cases, authors report a satisfying to excellent level of consistency among data sources. When ground truthing is difficult, as is often the case, other metrics could be developed to assess data robustness. We also strongly advocate cross-referencing results across multiple iEcology data sources, to test consistency of patterns [54,55]. Furthermore, culturomics can provide critical support to understand societal perceptions, interests, and values that affect the process of data generation [8,56].

Correct taxonomic identification in iEcology may be a cause for concern when compared with traditional ecological research. This may be true at several levels – from species misidentification by data producers to challenges that experts face when identifying species based on a limited number of images or videos of an individual organism. Furthermore, automated classification of species also generates misidentifications. Such embedded errors could also arise in other types of ecological data, such as life history traits, behavior, and abiotic variables. However, we expect that as iEcology sources increase in size, and methods to validate them improve, so will the ability to identify the extent and type of such problems in the data. Furthermore, we also suggest assigning a validity attribute to data that can be nonbinary, and dependent on the contributor's reputation and the likelihood of an observation – as is currently practiced on some citizen-science platforms.

iEcology research would greatly benefit from collaborative efforts and sharing of data, resources, and tools. These could be aided by developing specific metadata standards for sharing such data, which could include APIs and specific machine-learning algorithms used to extract or manipulate the data. Such developments could draw from similar efforts that are already being carried out by big ecological databases (e.g., www.gbif.org) to develop similar standards, which would make ecological data more interoperable [57].

iEcology repositories could be either centralized or remain decentralized, with benefits associated with both options. Centralized repositories would greatly aid the maintenance of high standards, as well as providing better reproducibility, open access, and versioning. Nevertheless, necessary effort on pre- and postprocessing of data and metadata for uploading, and generally rigid structure of a centralized repository may actually deter people involved in more local, small-scale explorations, and may ultimately hinder data sharing and collaboration. As iEcology is still in its infancy, there may be some advantages in its remaining decentralized and more flexible at this current stage. Yet, as the methods, tools, and associated data increase in breadth and scope, a move towards more collated, managed, and centralized repositories will become more natural and pertinent. Nevertheless, we advocate that good record keeping and maintaining high metadata standards is of particular importance to iEcology.

Other considerations of iEcology data sources involve interpretation and reproducibility. Some sources lack transparency in the way the considered data were produced and manipulated (e.g., search engines such as Google). Inability to publish raw data (as per provider guidelines) could also cause issues with scientific journal protocols that require making these available. Furthermore, some sources lack stability in data scope, underlying algorithms, and access options. These are inherent issues with many online sources. To alleviate these concerns, we advocate: (i) good record keeping of protocols for data access, handling, versioning, and analysis; (ii) harmonization of methods [58] and standardization of metadata; (iii) publishing (when possible) raw data in freely accessible and stable repositories together with associated scripts; (iv) use of open-source data and software; and (v) keeping up to date with methodologies developed in other relevant fields (e.g., computational sociology) for assessing and addressing such issues.

iEcology research may give rise to several ethical issues, pertaining to both people and nature. Data shared online, especially on social media platforms, sometimes include explicit personal information, while implicit information could also be used to identify individuals or to extract sensitive information. Therefore, the privacy of individuals and their identifiers should be maintained in both data repositories and iEcology outputs, adhering to the highest ethical standards [59]. Moreover, data sources that include precise information on locations and other key attributes of rare or endangered species could increase their exposure to poachers and collectors [60]. This threat could be alleviated by either restricting access to data on species deemed at risk, or limiting precision of open-access information. In general, servers holding iEcology data should be securely maintained to avoid such abuse.

## Concluding Remarks and Future Perspectives

The field of ecology is undergoing a rapid shift towards indirect, technology-based, and automated observations of nature and biodiversity [53,61,62], where iEcology is likely to play a critical role. Utilization of iEcology methods and data has greatly expanded recently, with most publications appearing during the past few years. iEcology is likely to experience rapid development over the coming years and become one of the major research techniques in ecology. While classical biodiversity research is irreplaceable for understanding the natural world through targeted observations and experiments, iEcology could provide novel and low-cost support for ongoing research efforts. The value of iEcology is likely to greatly increase as the global coverage of the Internet, mobile computing, sensor networks, and their users, expand. This will be augmented by leaps in computational capabilities, emergence of new data sources and types (such as odors, obtained by electronic noses) [63], and other emerging tools and technologies for using these data (Box 1). Combining these with other Big Data sources and efforts to understand nature, such as ecoinformatics, could also prove valuable.

In the near future, we foresee complete automation of all stages of data handling within iEcology, from access to visualization, to creation of ever-expanding datasets of biological entities, traits, behaviors, etc. This could give rise to a global digital monitoring initiative for the natural world. For example, an ecologist interested in animal behavior could produce tools to automatically scrape all uploaded YouTube videos for animal representations, automatically analyze them for different types of behaviors, include these into a constantly updated dataset, and ultimately analyze them in real time, to produce continuously updated research outputs. However, good expertise regarding the organisms studied and underlying ecological mechanisms at play will always be invaluable to make sense of these rapidly accumulating data and their inherent biases. Furthermore, many of the examples presented above demonstrate imagination and creativity in using data sources for ecology that were collected for other purposes. Above all, iEcology will benefit from such creativity to find new ways to harness data beyond their original purposes.

iEcology provides fertile ground for interdisciplinary collaborations, enhanced by a wide range of expertise and specializations. Furthermore, iEcology will create new opportunities for partnerships between academia, industry, governmental, and non-governmental organizations, working synergistically to produce original insights into the natural world.

## Outstanding Questions

How do iEcology insights differ from those uncovered by other ecological methods (in scope, reliability, applicability etc.)?

Can insights and observed patterns from regions with high digital coverage be generalized to those without?

How should we attach uncertainty to data from iEcology sources?

Should we aim for centralized or decentralized repositories and datasets?

Are there particular skills that should be taught to students to help them develop iEcology expertise and how can these be integrated in curricula?

Are Linnean or Wallacean shortfalls manifested in iEcology sources similar to classical ecological sources?

Will the future development of iEcology cause greater detachment from nature even among naturalists or ecologists and how could this be averted?

Will the use of iEcology alleviate some of the ethical concerns of handling animals, and will it give rise to new ethical concerns?

## Acknowledgments

I.J.'s work was supported by the J. E. Purkyně Fellowship of the Czech Academy of Sciences. F.C. was supported through the Invacost grants by the ANR, the AXA Research Fund Chair for Invasion Biology and Biodiversa AlienScenarios. R.A.C. acknowledges funding from the Helsinki Institute for Sustainability Science (HELSUS) to E.D.M. E.D.M. thanks the European Research Council (ERC) for funding under the European Union's Horizon 2020 research and innovation program (grant agreement #802933). A.T.S. acknowledges the EU Horizon 2020 research and innovation programme funding (project grant No. 677039), and ERDF/ESF funding (CZ.02.1.01/0.0/0.0/16_025/0007417). U.R. is supported by the Israel Science Foundation (grant No. 406/19). J.A.F. was supported by a research fellowship from Merton College and BBSRC (BB/S009752/1) and acknowledges funding from NERC (NE/S010335/1). We thank Łukasz Dylewski, Pieter De Frenne, Zuzanna A. Jagiello, Shelan S. Jeawak, Alison Johnston, Gabriella R. M. Leighton, Peter Mikula, and Justin G. Schuetz for sharing data and figures. The authors also thank Andrea Stephens and two anonymous reviewers for providing helpful comments and suggestions that improved the quality of the paper.

## Supplemental Information

Supplemental Information associated with this article can be found online at https://doi.org/10.1016/j.tree.2020.03.003.

## References

1. Castells, M. (1996) *The Information Age: Economy*, Blackwell, Society and Culture
2. Hampton, S.E. *et al.* (2013) Big data and the future of ecology. *Front. Ecol. Environ.* 11, 156–162
3. LaDeau, S.L. *et al.* (2017) The next decade of big data in ecosystem science. *Ecosystems* 20, 274–283
4. Michener, W.K. and Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol. Evol.* 27, 85–93
5. Ghermandi, A. and Sinclair, M. (2019) Passive crowdsourcing of social media in environmental research: a systematic map. *Glob. Environ. Chang.* 55, 36–47
6. Ekman, A. and Litton, J.E. (2007) New times, new needs; e-epidemiology. *Eur. J. Epidemiol.* 22, 285–292
7. Bohannon, J. (2011) Google Books, Wikipedia, and the future of culturomics. *Science* 331, e6395
8. Ladle, R.J. *et al.* (2016) Conservation culturomics. *Front. Ecol. Environ.* 14, 269–275
9. Di Minin, E. *et al.* (2015) Prospects and challenges for social media data in conservation science. *Front. Environ. Sci.* 3, 63
10. Sutherland, W.J. *et al.* (2018) A 2018 horizon scan of emerging issues for global conservation and biological diversity. *Trends Ecol. Evol.* 33, 47–58
11. Gaston, K.J. and Blackburn, T.M. (2000) *Pattern and Process in Macroecology*, Blackwell Science
12. Schuetz, J.G. and Johnston, A. (2019) Characterizing the cultural niches of North American birds. *Proc. Natl. Acad. Sci. U. S. A.* 116, 10868–10873
13. Barve, V. (2014) Discovering and developing primary biodiversity data from social networking sites: a novel approach. *Ecol. Inform.* 24, 194–199
14. Daume, S. (2016) Mining Twitter to monitor invasive alien species – an analytical framework and sample information topologies. *Ecol. Inform.* 31, 70–82
15. Dylewski, Ł. *et al.* (2017) Social media and scientific research are complementary –YouTube and shrikes as a case study. *Sci. Nat.* 104, 48
16. ElQadi, M.M. *et al.* (2017) Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. *Ecol. Inform.* 39, 23–31
17. Hong, S. *et al.* (2017) Conservation activities for the Eurasian otter (*Lutra lutra*) in South Korea traced from newspapers during 1962–2010. *Biol. Conserv.* 210, 157–162
18. Jeawak, S.S. *et al.* (2017) Using Flickr for characterizing the environment: an exploratory analysis. In *13th International Conference on Spatial Information Theory (COSIT 2017)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik
19. Jeawak, S.S. *et al.* (2018) Mapping wildlife species distribution with social media: Augmenting text classification with species names. In *10th International Conference on Geographic Information Science (GIScience 2018)* (Winter, S. *et al.*, eds), pp. 45:1–45:6
20. Hart, A.G. *et al.* (2018) Testing the potential of Twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa. *Methods Ecol. Evol.* 9, 2194–2205
21. Allain, S.J. (2019) Mining Flickr: a method for expanding the known distribution of invasive species. *Herpetol. Bull.* 148, 11–14
22. Fukano, Y. and Soga, M. (2019) Spatio-temporal dynamics and drivers of public interest in invasive alien species. *Biol. Invas.* 21, 3521–3532
23. Pace, D.S. *et al.* (2019) An integrated approach for cetacean knowledge and conservation in the central Mediterranean Sea using research and social media data sources. *Aquat. Conserv.* 29, 1302–1323
24. Giovos, I. *et al.* (2016) Social media in the service of conservation: a case study of dolphins in the Hellenic seas. *Aquat. Mamm.* 42, 12–19
25. Jiménez-Valverde, A. *et al.* (2019) Photo-sharing platforms key for characterising niche and distribution in poorly studied taxa. *Insect Conserv. Divers.* 12, 389–403
26. Hentati-Sundberg, J. and Olsson, O. (2016) Amateur photographs reveal population history of a colonial seabird. *Curr. Biol.* 26, R226–R228
27. De Frenne, P. *et al.* (2018) Using archived television video footage to quantify phenology responses to climate change. *Methods Ecol. Evol.* 9, 1874–1882
28. Foglio, M. *et al.* (2019) Animal wildlife population estimation using social media images collections. *arXiv preprint*, 1908.01875
29. Francis, F.T. *et al.* (2019) Shifting headlines? Size trends of newsworthy fishes. *PeerJ* 7, e6395
30. Jiménez-Alvarado, D. *et al.* (2019) Historical photographs of captures of recreational fishers indicate overexploitation of nearshore resources at an oceanic island. *J. Fish Biol.* 94, 857–864
31. Breckheimer, I.K. *et al.* (2020) Crowd-sourced data reveal social–ecological mismatches in phenology driven by climate. *Front. Ecol. Environ.* 18, 76–82
32. Mittermeier, J.C. *et al.* (2019) A season for all things: phenological imprints in Wikipedia usage and their relevance to conservation. *PLoS Biol.* 17, e3000146
33. Gonella, P.M. *et al.* (2015) *Drosera magnifica* (Droseraceae): the largest New World sundew, discovered on Facebook. *Phytotaxa* 220, 257–267
34. Rahayu, S. and Rodda, M. (2019) *Hoya amicabilis* sp. nov. (Apocynaceae, Asclepiadoideae), from Java discovered on Facebook. *Nord. J. Bot.* 37, e02563
35. Leighton, G.R. *et al.* (2016) Just Google it: assessing the use of Google Images to describe geographical variation in visible traits of organisms. *Methods Ecol. Evol.* 7, 1060–1070
36. Miranda, E.B. *et al.* (2016) The ecology of human-anaconda conflict: a study using internet videos. *Trop. Conserv. Sci.* 9, 43–77

37. Mikula, P. *et al.* (2018) Large-scale assessment of commensalistic–mutualistic associations between African birds and herbivorous mammals using internet photos. *PeerJ* 6, e4520
38. Jagiello, Z.A. *et al.* (2019) What can we learn about the behaviour of red and grey squirrels from YouTube? *Ecol. Inform.* 51, 52–60
39. Fisher, J. and Hinde, R.A. (1949) The opening of milk bottles by birds. *Br. Birds* 42, 347–357
40. Gil, M.A. *et al.* (2018) Social iformation links individual behavior to population and community dynamics. *Trends Ecol. Evol.* 33, 535–548
41. Firth, J.A. (2020) Considering complexity: animal social networks and behavioural contagions. *Trends Ecol. Evol.* 35, 100–104
42. Elmer, F. *et al.* (2019) Black spot syndrome in reef fishes: using archival imagery and field surveys to characterize spatial and temporal distribution in the Caribbean. *Coral Reefs* 38, 1303–1315
43. Haas, A.F. *et al.* (2015) Can we measure beauty? Computational evaluation of coral reef aesthetics. *PeerJ* 3, e1390
44. Becken, S. *et al.* (2019) A hybrid is born: integrating collective sensing, citizen science and professional monitoring of the environment. *Ecol. Inform.* 52, 35–45
45. Proulx, R. *et al.* (2014) Googling trends in conservation biology. *Conserv. Biol.* 28, 44–51
46. Snijders, L. *et al.* (2017) Animal social networks can help wildlife conservation. *Trends Ecol. Evol.* 32, 567–577
47. Brakes, P. *et al.* (2019) Animal cultures matter for conservation. *Science* 363, 1032–1034
48. Sullivan, M. *et al.* (2019) Social media as a data resource for monkseal conservation. *PLoS One* 14, e0222627
49. Chamberlain, J. (2018) Using social media for biomonitoring: how Facebook, Twitter, Flickr and other social networking platforms can provide large-scale biodiversity data. *Adv. Ecol. Res.* 59, 133–168
50. Bollier, D. and Firestone, C.M. (2010) *The Promise and Peril of Big Data*, Aspen Institute
51. Christin, S. *et al.* (2019) Applications for deep learning in ecology. *Methods Ecol. Evol.* 10, 1632–1644
52. Ladle, R.J. *et al.* (2019) A culturomics approach to quantifying the salience of species on the global internet. *People Nat.* 1, 524–532
53. Pimm, S.L. *et al.* (2015) Emerging technologies to conserve biodiversity. *Trends Ecol. Evol.* 30, 685–696
54. Hausmann, A. *et al.* (2018) Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conserv. Lett.* 11, e12343
55. Correia, R.A. *et al.* (2019) Inferring public interest from search engine data requires caution. *Front. Ecol. Environ.* 17, 254–255
56. Gaston, K.J. *et al.* (2018) Personalised ecology. *Trends Ecol. Evol.* 33, 916–925
57. Wieczorek, J. *et al.* (2012) Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One* 7, e29715
58. Musche, M. *et al.* (2019) Research questions to facilitate the future development of European long-term ecosystem research infrastructures: a horizon scanning exercise. *J. Environ. Manag.* 250, 109479
59. Monkman, G.G. *et al.* (2018) The ethics of using social media in fisheries research. *Rev. Fish. Sci. Aquac.* 26, 235–242
60. Lindenmayer, D. and Scheele, B. (2017) Do not publish. *Science* 356, 800–801
61. Bohan, D.A. *et al.* (2017) Next-generation global biomonitoring: large-scale, automated reconstruction of ecological networks. *Trends Ecol. Evol.* 32, 477–487
62. Kitzes, J. and Schricker, L. (2019) The necessity, promise and challenge of automated biodiversity surveys. *Environ. Conserv.* 46, 247–250
63. Sutherland, W.J. *et al.* (2017) A 2017 horizon scan of emerging issues for global conservation and biological diversity. *Trends Ecol. Evol.* 32, 31–40
64. Jepson, P. and Ladle, R.J. (2015) Nature apps: waiting for the revolution. *Ambio* 44, 827–832
65. Buettel, J.C. and Brook, B.W. (2016) Egress! How technophilia can reinforce biophilia to improve ecological restoration. *Restor. Ecol.* 24, 843–847
66. Dorward, L.J. *et al.* (2017) Pokémon Go: benefits, costs, and lessons for the conservation movement. *Conserv. Lett.* 10, 160–165
67. Di Minin, E. *et al.* (2018) Machine learning for tracking illegal wildlife trade on social media. *Nat. Ecol. Evol.* 2, 406–407
68. Norouzzadeh, M.S. *et al.* (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5716–E5725
69. Kaur, K.M. *et al.* (2019) Using text-mined trait data to test for cooperate-and-radiate co-evolution between ants and plants. *PLoS Comput. Biol.* 15, e1007323
70. Di Minin, E. *et al.* (2019) A framework for investigating illegal wildlife trade on social media with machine learning. *Conserv. Biol.* 33, 210–213
71. Wearn, O.R. *et al.* (2019) Responsible AI for conservation. *Nat. Mach. Intell.* 1, 72–73
72. Aide, T.M. *et al.* (2013) Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103
73. Harris, S.A. *et al.* (2016) Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods Ecol. Evol.* 7, 713–724
74. Linke, S. *et al.* (2018) Freshwater ecoacoustics as a tool for continuous ecosystem monitoring. *Front. Ecol. Environ.* 16, 231–238
75. Rajan, S.C. *et al.* (2019) Rapid assessment of biodiversity using acoustic indices. *Biodivers. Conserv.* 28, 2371–2383
76. Firdaus, A. *et al.* (2019) The rise of "blockchain": bibliometric analysis of blockchain study. *Scientometrics* 120, 1289–1331
77. Atzori, L. *et al.* (2010) The internet of things: A survey. *Comput. Netw.* 54, 2787–2805
78. Berger-Tal, O. and Lahoz-Monfort, J.J. (2018) Conservation technology: the next generation. *Conserv. Lett.* 11, e12458
79. Hill, A.P. *et al.* (2019) Leveraging conservation action with open-source hardware. *Conserv. Lett.* 12, e12661
80. Galaz, V. *et al.* (2010) Can web crawlers revolutionize ecological monitoring? *Front. Ecol. Environ.* 8, 99–104
81. Recknagel, F. (2008) Ecological informatics: Overview. In *Encyclopedia of Ecology* (Vol. 2) (Jorgensen, S.E. and Faith, B.D., eds), pp. 1041–1058, Elsevier