Kent Academic Repository

Thompson, Ruth M., Hall, Jordan, Morrison, Chris, Palmer, Nicole R. and Roberts, David L. (2021) *Ethics and governance for internet-based conservation science research.* Conservation Biology . ISSN 0888-8892.

Ethics and governance for internet-based conservation science research

Ruth M Thompson[1], Jordan Hall[2], Chris Morrison[3], Nicole R Palmer[4], David L Roberts[1,5,6]

[1] Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, Marlowe Building, University of Kent, Canterbury, Kent, CT2 7NR U.K.

[2] Information Compliance Office, Darwin College, University of Kent, Canterbury, Kent, CT2 7NY U.K.

[3] Copyright, Licensing & Policy, Information Services, Templeman Library, University of Kent, Canterbury, Kent, CT2 7NU U.K.

[4] Research Ethics and Governance, Research Services, The Registry, University of Kent, Canterbury, Kent, CT2 7NZ U.K.

[5] Department of Zoology, University of Oxford, Oxford, OX1 3SZ U.K.

[6] Oxford Martin School, University of Oxford, 34 Broad Street, Oxford, OX1 3BD U.K.

Corresponding Author:

Ruth M Thompson
rmt25@kent.ac.uk
Durrell Institute of Conservation and Ecology
School of Anthropology and Conservation
Marlowe Building
University of Kent
Canterbury
Kent, UK
CT2 7NR

Article impact statement: Conservationists undertaking internet-based research should do more to avoid ethical and legal

backlash from users and service providers.

Running head:

Cyberethics

Keywords:

compliance, copyright, cyberethics, data protection, integrity, intellectual property, internet mediated, privacy

Abstract

Internet-based research is increasingly important for conservation science with wide-ranging applications and contexts, including culturomics, the illegal wildlife trade, and citizen science. However, online research methods pose a range of ethical and legal challenges. Online data may be protected by copyright, database rights, or contract law. Privacy rights may also restrict the use and access of data, as well as ethical requirements from institutions. Online data has real-world meaning, and the ethical treatment of individuals and communities must not be marginalized when conducting Internet-based research. As ethics frameworks originally developed for biomedical applications are inadequate for these methods, we propose that research activities involving the analysis of pre-existing online data is treated analogously to offline social science methods, in particular, non-deceptive covert observation. By treating Internet users and their data with respect and due consideration, conservationists can uphold the public trust needed to effectively address real-world issues.

Introduction

Internet-based research is increasingly being utilized in conservation science across many methodological contexts, including analyzing trends in public interest through site traffic and search engine usage (Soriano-Redondo et al. 2017; Fernández- Bellon & Kane 2019), studying online trade in wildlife products through manual searches or automated retrieval (Harrison et al. 2016; Sung & Fong 2018), behavior change (Doughty et al. 2020), outreach and citizen science projects (Tulloch et al. 2013), and the sharing of databases and analytical tools online. Conservation culturomics, which analyzes trends in word usage over the Web (such as search engine queries of endangered species), is also emerging as a major field, requiring vast quantities of online data (Sutherland et al. 2018). Online research activities centered on collecting and sharing data relating to or created by individuals requires consideration of research ethics, intellectual property (including copyright and database rights), privacy rights and data protection (Fig. 1) (Franzke et al. 2020). Ethical Internet-based research is a particular challenge for conservationists operating in a normative, crisis discipline reliant on support from stakeholders and the wider public, as well as navigating potential social imbalances or mixed roles of the researcher in relation to study subjects (Bennett et al. 2017; Brittain et al. 2020).

Having a clear and explicit rationale behind online research activities from the start upholds public accountability and confidence when disseminating results (Zook et al. 2017; Monkman et al. 2018). The COVID-19 pandemic and resulting social distancing measures have also led to otherwise offline research activities, including interviews or surveys, being conducted online. This change in operations requires researchers to adapt their practices, ensuring their use of technology does not conflict with participants' ethical expectations.

In this paper, we focus on the ethics and legal implications of online information access and collection, and the appropriate dissemination and sharing of data and analytical results. We start with the foundation of free and prior informed consent in biomedical ethics, and discuss its feasibility for Internet research depending on how the data of interest is created. We argue that justifying research through definitions of privacy or the application of social norms is not appropriate in the absence of consent, and instead argue that the collection of pre-existing data should be facilitated through non-deceptive covert observation where possible (Spicker 2011) and remain relevant to the contexts in which the data is created (Nissenbaum 2010). Organizational codes of conduct and publication requirements should be updated to reflect ethical challenges specific to Internet-based research for conservation science. Finally, we discuss legal challenges and risks to researchers in terms of intellectual property, contract law, privacy and data protection. In addition to the topics discussed here, researchers will need to assess the ethical implications of analytical methods used, such as the interpretation of machine learning tools and its role in predictions or decision-making (Wearn et al. 2019). Long-term storage and data sharing are also becoming important elements in funding and publication requirements, with implications for how research is planned to account for the relevant consent or permissions needed, and constructing datasets with privacy built-in (Hart et al. 2016). For a multidisciplinary perspective, conservationists should refer to guidelines by the Association of Internet Researchers (AoIR) (Ess et al. 2002; Markham & Buchanan 2012; Franzke et al. 2020). The paper presented here does not constitute legal advice.

Ethics and contextual integrity in Internet-based research

Social science methods have become increasingly prominent in conservation with the understanding that anthropogenic biodiversity and habitat loss cannot be countered with a purely natural sciences focus (Bennett et al. 2017; Brittain et al. 2020). The ethical underpinnings of social research for conservation goals using methods preceding the Internet, such as surveys and interviews, are largely guided by codes of conduct established in other disciplines; in turn, many of these are rooted in biomedical ethics (Ibbett & Brittain 2019). The Belmont Report summarizes these ethical standards as ideals of respect (preserving autonomy and protecting those with less autonomy), beneficence (minimizing harm, benefiting individuals and society) and justice (fairness) (National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research 1979). These ideals are typically upheld through free and prior informed consent (FPIC) of research participants with the option of withdrawal during the study, the assurance of confidentiality or anonymity, and the validation of ethical review. Ethical approval is typically conducted once prior to the start of a study, and forms a more rigid, 'top-down' regulatory approach necessary in medical or psychological experimentation. Methods suited to this ethical process have pre-determined, specific variables and study subjects, unlike commonly used Big Data methods in Internet research such as pattern discovery over a broad range of attributes.

Internet-based research can be broadly categorized into two areas based on whether the original data of interest is created for the purpose of the research (e.g. respondents completing an online survey or answering interview questions created by the researcher) or if the data is created by an Internet user for their own purposes (e.g. posting to an online traditional medicine forum, listing an exotic pet for sale, or sending a query on reintroduced species to a search engine). For the former, the use of Internet-mediated communication with participants is analogous to offline methods that require FPIC. There have been increasing calls within conservation for a process-oriented, reflective ethics approach to social science methods used in fieldwork requiring awareness of value conflicts, power dynamics and cultural sensitivity (St John et al. 2016; Ibbett & Brittain 2019; Brittain et al. 2020). In addition to these recommendations, researchers must ensure their use of technology does not undermine ethical principles (e.g., assessing and disclosing the privacy and security risks of conducting surveys through third party websites to participants, and encrypting sensitive emails to maintain confidentiality (Franzke et al. 2020, p.80)).

For research involving the collection and analysis of pre-existing online data, within legal constraints, the situation becomes more contentious. Even if researchers view this information as ‘fair game ’ by considering all accessible online data as authored works or activities consciously under public scrutiny, Internet users’ expectations of privacy often extend to publicly accessible data or information that is contractually accessible by third parties (Boyd & Crawford 2012). Furthermore, given the rapid growth of the online world and its encroachment into all aspects of many people's lives, non-participation in online activities can result in social exclusion, economic suppression and incomplete health, governmental and other administrative services (Nissenbaum 2011). As such, it is no longer appropriate to assume informed, non-coercive consent. However, FPIC is most often impractical for online research requiring large and/or complete datasets of pre-existing information for meaningful analysis (particularly when studying illegal activity), and explicit opt-in/opt-out consent requests can itself be viewed as intrusive behavior (Ess et al. 2002, Spicker 2011). Given that public trust and cooperation are necessary for positive conservation intervention outcomes, being able to provide transparent, ethical reasoning for Internet research methods is still important, even if the information used is freely available.

The ideals on which FPIC is based - respect, beneficence and justice - in research utilizing pre-existing data are closely tied with peoples’ notions of privacy and violations thereof. Privacy is beneficial to individuals and societies through preserving autonomy and freedoms, while shielding people from unjust treatment (Nissenbaum 2010). Despite its importance, however, assigning a strict universal definition to privacy in order to build an ethical framework is not straightforward; in short, it can hold a multitude of potentially contradictory meanings and is contextually dependent (Ibid.). In attempting to circumvent the aforementioned issues in obtaining FPIC, the collection and analysis of pre-existing data has been justified by definitions of public spaces as a negation of privacy rights, or by downgrading peoples' online representations as digital subjects resulting in dehumanization through abstraction (Buchanan 2017). Regulatory approaches to privacy typically build on binary classifications of private and public spaces, or sensitive and non-sensitive personal details, with corresponding restrictive or permissive approaches to the accessibility or use of information (Nissenbaum 2010). However, as illustrated above, these regulatory definitions do not necessarily align with users' expectations or principles. Another viewpoint of conducting research that demonstrates respect, beneficence and justice is the need to recognize and work within the contexts and boundaries of subjects' values and social norms. For other social science techniques employed in conservation, such as interviews, this is straightforward: the research is conducted within specific contexts with a limited number of participants, and the principles and methodology of the study are explicitly agreed to through FPIC. FPIC, in this sense, provides a mutually understood ethical contract between the researcher and research subject. However, for online research methods where collected data may relate to anyone, anywhere, relying instead on universal norms as perceived and formulated by the researcher or discipline as a form of assumed consent becomes problematic. Even if the researcher believes that their actions are justified, this could translate differently to others in a global context as paternalistic or have unforeseen consequences. In addition, the use of normative constructs (words or phrases which carry some implicit judgment or value) in decision-making processes for conservation has been criticized for lacking scientific or logical rigor (Yanco et al. 2019). This highlights the need to clearly define conservationists' underlying value assumptions when formulating a research plan as well as interventions. These complications become amplified as the proposed benefits can be focused on non-human species and may come at a cost to human groups, often separate from the researchers themselves (Ibbett & Brittain 2019; c.f. Newing & Perram 2019). Both of these instances (justifications based on definitions of privacy or universal application of perceived social norms), can lead to public disapproval and a setback in conservation efforts. Public trust is not guaranteed simply by following the minimum legal requirements or formulating abstract moral justifications from the researchers' perspective.

In response to public unease with technology replacing or altering the means by which social interaction, transactions or the provision of services take place, Nissenbaum's theory of contextual integrity (Nissenbaum 2010) argues that instead of attempting to define privacy and violations thereof, the flow of information should be described in terms of the contexts and sub-contexts in which it operates (e.g. commerce, education), persons or entities involved (e.g. sellers, platforms), the nature of the information and the principles by which it is transmitted (e.g. compulsory, confidential). If the new flow of information departs from entrenched norms specific to those contexts and has harmful implications to individuals, social structures or the functional purposes and goals of the context, then this is in violation of contextual integrity and should be avoided (Nissenbaum 2010, p.181-183). While this theory is a useful means of conceptualizing research data in context, collecting and analyzing this for conservation is (unless, for example, under the mandate of legal monitoring) not functionally relevant to the information flow taking place for the sender or intended recipients. How, then, can conservationists appropriately act within this ecosystem of data?

Covert research ethics

One approach to maintaining ethical integrity when research involves pre-existing data could be from considering these activities as observation analogous to its offline counterpart, that is, covert research methods in the social sciences. Spicker (2011) defines covert research as consisting of non- or limited disclosure to research subjects, as opposed to being intrinsically linked with deceptive practices or avoidance of detection (e.g. Roulet et al. 2017). For example, studying publicly available information on exotic pets for sale online would be covert observation; the sellers are not explicitly informed that the data they upload is also being used for research (e.g. Sung & Fong 2018). However, if this information is in a closed social media group requiring approval under a provided identity, posing as a potential buyer to gain access would be an act of deceptive covert research. Unless such exceptional practices are rigorously justified by carefully balancing competing principles (Spicker 2011), the researcher should make their purposes explicit to gatekeepers and participants (e.g. Hinsley et al. 2016). It should be noted that research in a public space could still be viewed as deceptive, if the researcher knowingly exploits subjects' assumptions of their identity and purposes (Spicker 2011). We define 'public space' as data which can be physically accessed by anyone with the appropriate legal technological means and according to the sites' terms of service, rather than as the antithesis of a 'private space'. Since justifications based on researchers' intentions may be questionable or unclear, Nissenbaum's theory of contextual integrity (Nissenbaum 2010) can help determine the appropriate flow and use of information within the resulting multiplicity of contexts, such as the platform, community, or activity in which users engage. In this case, the appropriate flow of information ensures that data collection and processing remain applicable to the contexts in which the data has been created - in the above scenarios, a typical e-commerce website would be subject to legal conditions and the companies' terms of service for the appropriate sale of items. This limits the extent to which information can be reasonably collected and processed. For example, if studying online marketplaces that may host illegal wildlife trade, aggregating this data with sellers' social media profiles on a separate site would be ethically dubious, especially if considering such a profile as being a composite of personal information and that of uninvolved family and friends. However, comparing sale items with official import records (Sung & Fong 2018) would be contextually relevant and not obviously objectionable.

Complying with prevailing contextual norms is not equivalent to assuming consent: it can be argued that contextual integrity leads to a "tyranny of the normal", that is, a conservative framework with adherence to norms that reflect a majority of users at the potential expense of minority groups or individuals (Nissenbaum 2010, p.160). Researchers should recognize and minimize any potential harms to individuals and groups who may find the methods disagreeable without the researchers' knowledge. This includes de-identifying, anonymizing, summarizing and/or minimizing collected data sets as soon as possible, evaluating the possible broader consequences of sharing research results and data other than intended conservation outcomes (Di Minin et al. 2021), and re-evaluating the contexts in which researchers operate in light of their observations, including the exclusion of groups found to oppose the collection of data where reasonably possible. A key difference between conservation science and many other research activities is that it is operating as a crisis discipline, and whether or not conservationists are acting in a proactive or reactive capacity, the observations they make online are not for passive understanding but for monitoring and intervention, with implications for individuals' and communities' autonomy. Defining these research methods as covert observation recognizes that prior informed consent does not take place, and this burden of responsibility must be acknowledged even if the activity involved seems trivial to the researcher.

There is no single perspective or unifying framework in which to conduct online research without some consequence or risk. Online conservation research will frequently result in situations for which there is no single correct answer or ethical shortcut. Researchers may be tempted to “cherry-pick” abstract moral theories that suit their preferred methods, instead of critically reflecting on the real-world circumstances of their work (Macfarlane 2009, p.154). The AoIR recommend that through practical experience, familiarity with a range of ethical frameworks and discussion with peers and stakeholders, researchers can develop sound judgement that reflects the flexibility required to adapt to novel research contexts (Franzke et al. 2020).

Conservation research in commercial contexts

Since the commercialization of the Internet in the 1990s (Langford 2000, p.33), and the subsequent domination of commercial platforms on the World Wide Web, the data that researchers obtain from and about users are usually mediated by corporate entities with separate business interests which may conflict with those of users and conservationists (Toivonen et al. 2019). As such, the resulting information may be fragmented to protect commercial interests (Ladle et al. 2016), distorted by user expectations akin to an “informational panopticon” leading to self-censorship or editing (Nissenbaum 2010, p.75), or the data may be molded for commercial use, limiting its meaning. Even if users would opt for services with greater privacy controls, the terms of service or other legal agreements they are subject to under the “duty to read" are often classified as unreadable to the average US citizen (Benoliel & Becher 2019), reducing market competition and contributing to user disempowerment, particularly among vulnerable groups. Researchers should be aware of these contexts to better assess the exploitation of subjects involved.

Publication requirements for ethical research

Since social science methods preceding Internet-based research have been utilized for conservation science, such as interviews, publishers in the field have frequently retained ethical requirements associated with biomedical research such as the Helsinki declaration (WMA General Assembly 2001), or indirectly through approval from ethics committees (or their equivalent) that retain similar policies (Ibbett & Brittain 2019). Ibbett and Brittain (2019) also reported that, in the context of wildlife hunting, while there is an upward trend in reporting ethical considerations in conservation publications with social science methods, there is a disparity between journal requirements and their implementation. St. John et al. (2016) recommend that the American Anthropological Association (AAA) Code of Ethics (2012) is more suitable for individualized human subject interaction, although how applicable this is to research using subjects' pre-existing online data is uncertain. In its 2012 revision, the AAA's Principles of Professional Responsibility state that research without prior or retroactive informed consent should be avoided (see also the Society for Conservation Biology (2004)), suggesting an opposition to a significant portion of online research methods for large-scale, publicly accessible data. While analysis of anonymized data would not require individuals' consent according to these guidelines, seemingly innocuous online data can frequently be aggregated with other resources for the re-identification of individuals on a scale and level of detail that surpasses offline research (Tavani & Grodzinsky 2019). Instead of interpreting guidelines from other disciplines, updated ethical requirements from publishers specific to the challenges of conservation research and online research methods could contribute to more cohesive ethical practices and reporting. This would be especially valuable to conservationists who do not have access to external validation mechanisms such as ethical review boards (Ibbett & Brittain 2019).

Compliance and legal risks in online research

Unlike most offline research activities, data collected via the Web are typically owned by companies or individuals, with corresponding legal protections, and transmitted by commercial entities. In addition to maintaining ethical integrity, conservationists' online research activities must also navigate these areas of compliance and assess the risks where legal uncertainty arises. Compliance with the law is a first indicator to the public as to whether researchers are acting ethically, and is especially relevant if conservationists seek to influence legislation through their results.

Legal compliance in Internet-based research is generally centered on copyright, privacy and data protection. Copyright protection is broadly recognized internationally, with variations in implementation and exceptions between member countries (World Intellectual Property Organization 1979; World Trade Organization 1994), whereas privacy rights and data protection are more varied and largely dictated by pre-existing social norms (Spinello 2017, p.173).

Intellectual property and copyright

Digital media, whether created by individuals or companies, is automatically protected under international copyright laws which confer economic (financial reward) and moral (personal credit/reputation) rights to the work's owner (World Intellectual Property Organization 1979). These protections are still relevant to media published online, including images, text, audio, code and compilations of data (World Trade Organization 1994; Ricketson 2003). Some countries and territories, such as within the EU, also confer database (*sui generis*) rights, which may include social media and auction sites' data structures (European Commission 2015).

Exceptions to copyright enable the copying and sharing of works without authorization, within a defined scope, and for public benefit, including non-profit research activities. Such exceptions are especially important in large-scale studies, where the ability to credit and gain permissions from content creators becomes impractical. Whilst international agreements such as the Berne Convention (World Intellectual Property Organization 1979) and the TRIPS agreement (World Trade Organization 1994 section 1 article 10) set out broad definitions of works protected by copyright and exclusive rights for authors and producers, the interpretation and implementation of copyright and its exceptions are defined by individual nation states. Text and data mining (TDM) exceptions have been gaining traction in the Global North for non-commercial research, with varying requirements. For example, the UK has had a TDM exception since 2014 (UK Public General Acts 1988) which requires a "sufficient acknowledgement" of the author or copyright owner, while the EU Digital Single Market Directive, to be implemented by member states within two years of approval, specifies that mandatory TDM exceptions will apply to research organizations whose activity is structurally non-commercial and in the public interest (i.e., the output is not subject to “preferential access” and does not “conflict with the normal exploitation of the works”), with some provisions for public-private partnerships and commercial use (European Parliament & Council of the European Union 2019). In the US, TDM exceptions are not explicitly defined and may fall under “fair use” on a case-by-case basis (Copyright Law of the United States (Title 17) 1976). TDM exceptions have been further extended to commercial activities in several instances, potentially enabling joint conservation research with the private sector; for example, Japan has had broad TDM exceptions to copyright since 2009 (Oyama et al. (Translation) 1970), and Singapore is expanding TDM exceptions with analytical requirements (Singapore Ministry of Law & Intellectual Property Office of Singapore 2019). In all cases, the researcher must have the legal means to access the works in question, and international or public-private collaborations should have the appropriate regional rights to copy, distribute or analyze data. Although TDM exceptions are gaining traction internationally, intellectual property law is often more restrictive in developing countries (Okediji 2019), and the current variations with which they are implemented or interpreted in a global context make them difficult to solely rely on for research purposes.

Even if intellectual property rights challenges can be overcome, data access may be restricted by contract law or technical measures. For example, if a researcher intends to scrape data from an online auction site, their lawful access will be subject to the service’s terms and conditions, regardless of whether registration is required. The terms of service are a binding legal contract between the user and service, but are rarely written to accommodate academic researchers, normally assuming that a consumer or rival commercial organization is accessing the platform. The permissions and prohibitions set out in these terms therefore require interpretation, potentially introducing uncertainty over and above intellectual property concerns. Access to data may also be restricted through technical measures such as digital rights management software. Legislation including the Digital Millennium Copyright Act (DMCA) in the US and article 6 of the 2001 Information Society Directive in the EU (European Parliament & Council of the European Union 2001) prevent circumvention of technical measures such that even if a research exception to copyright applies, bypassing these measures is legally ambiguous or practically difficult (Liu 2003; Spinello 2017, p.119). Given the legal uncertainties and regional variations presented by intellectual property and contract law, researchers should seek expert advice where possible.

Privacy and data protection

In contrast to somewhat harmonized approaches to copyright, there is much greater regional variation in approaches to privacy and data protection, which are driven by pre-existing social norms. As a starting point, the US Department for Health and Human Services maintains an international compilation of laws governing human research (Office for Human Research Protections 2020). The US and EU provide an example of how stark these contrasts in legislation can be. Privacy in US legislature is highly fragmented and focuses on privacy from the state, with the assumption that market self-regulation will ensure that users migrate to services that treat their data appropriately (Benoliel & Becher 2019). Dealing with data in this context requires knowledge of specific legislation related to the researchers' methods, data, and region of interest due to additional variability in data processing requirements between states. In the EU, privacy is regarded as a distinct human right with the General Data Protection Regulation (GDPR) imposing stringent requirements for organizations storing, processing or sharing citizens' data, with interpretation at the state level and extra-territorial applicability (European Parliament & Council of the European Union 2016; Spinello 2017, p.176). These requirements apply both to pre-existing online data collection and citizen science projects. It should be noted that ensuring ethical practices may require going beyond basic legal requirements to ensure fairness, for example, by implementing the same privacy protections for participants in a research study regardless of nationality.

In our brief outline of intellectual property, privacy and data protection laws, it becomes clear that the aims, methods, analysis and sharing of data with their respective legal implications must be considered throughout the research process. This may include input from organizational copyright and IP specialists, data protection officers and independent legal advice when necessary.

Conclusions

Conservation scientists are utilizing Internet-based research for a wide range of purposes from culturomics to the online wildlife trade, with increasing relevance and emerging applications (Sutherland et al. 2018). To benefit from these methods, the ethical and legal implications of using online data must be evaluated, such as privacy, the appropriate ethical treatment of Internet users, intellectual property and data protection requirements, along with the particular challenges these pose in conservation research contexts.

Free and prior informed consent is often impractical when collecting users' pre-existing online data for research. Despite this, it is important to recognize the ethical implications of these activities as research involving human subjects. By treating these online methods analogously to covert research, in particular non-deceptive covert observation wherever possible (Spicker 2011), these activities can be assessed honestly as a balance of competing principles within the contexts they operate in on a case-by-case basis. Consent cannot be assumed on the part of individuals regardless of social norms that represent the views of a majority within such contexts (Nissenbaum 2010, p.160; Markham & Buchanan 2012). Potential resulting harms should be minimized regardless of data anonymization, bearing in mind that innocuous combinations of users' non-personal information (e.g. details of online wildlife sales associated with a user) can still lead to re-identification when publishing results or sharing datasets. Researchers should avoid treating ethics as an administrative checklist, as being unfamiliar with the underlying reasoning or decision-making processes for ethical research practice can leave them unprepared for novel situations (Geller et al. 2010).

Without awareness of the scope of copyright and contract law in their region, researchers could breach these laws from the use of online images, text, or audio. Separate database rights can also be extended to online structured data including e-commerce sites and social media. Conservationists should be aware that non-compliance with copyright and contract law can lead to personal and organizational legal risk, as well as potentially damaging public relations if they act illegally.

By carefully considering the ethical treatment of Internet users, in addition to complying with regulations, conservation scientists can uphold the public trust needed for successful intervention outcomes to pressing issues.

Acknowledgments:

We thank Rosalyn Bass for her insight on the implications of long-term data storage and access on ethical research planning.

References

American Anthropological Association. 2012. AAA Statement on Ethics: Principles of Professional Responsibility. Available from https://www.americananthro.org/LearnAndTeach/Content.aspx?ItemNumber=22869 (accessed December 2020).

Bennett NJ et al. 2017. An appeal for a code of conduct for marine conservation. Marine Policy **81**:411–418 DOI: 10.1016/j.marpol.2017.03.035.

Benoliel U, Becher SI. 2019. The Duty to Read the Unreadable. Boston College Law Review **60**(8):2255-2296 DOI: 10.2139/ssrn.3313837.

Boyd D, Crawford K. 2012. Critical Questions For Big Data. Information, Communication & Society **15**:662-679 DOI: 10.1080/1369118X.2012.678878.

Brittain S et al. 2020. Ethical considerations when conservation research involves people. Conservation Biology **34**:925-933 DOI: 10.1111/cobi.13464.

Buchanan E. 2017. Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. PLOS ONE **12**(12):e0187155 DOI: 10.1371/journal.pone.0187155.

Copyright Law of the United States (Title 17). 1976. Available from https://www.copyright.gov/title17/92chap1.html (accessed December 2020).

Di Minin E, Fink C, Hausmann A, Kremer J, Kulkarni R. 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology **35**:437-446 DOI: 10.1111/cobi.13708.

Doughty H, Wright J, Veríssimo D, Lee JSH, Oliver K, Milner-Gulland EJ. 2020. Strategic advertising of online news articles as an intervention to influence wildlife product consumers. Conservation Science and Practice **2**. DOI: 10.1111/csp2.272.

Ess C, the AoIR ethics working committee. 2002. Ethical decision-making and Internet research: Recommendations by the AoIR ethics working committee. Association of Internet Researchers. Available from www.aoir.org/reports/ethics.pdf (accessed December 2020).

European Commission. 2015, April 25. Evaluation of Directive 96/9/EC on the legal protection of databases. Available from https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection (accessed December 2020).

European Parliament, Council of the European Union. 2001. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. Official Journal of the European Union. Available from https://eur-lex.europa.eu/eli/dir/2001/29/oj (accessed December 2020).

European Parliament, Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. Available from https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed December 2020).

European Parliament, Council of the European Union. 2019. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. Official Journal of the European Union. Available from https://eur-lex.europa.eu/eli/dir/2019/790/oj (accessed December 2020).

Fernández- Bellon D, Kane A. 2019. Natural history films raise species awareness—A big data approach. Conservation Letters e12678 DOI: 10.1111/conl.12678.

Franzke AS, Bechmann A, Zimmer M, Ess C, Association of Internet Researchers. 2020. Internet Research: Ethical Guidelines 3.0. Available from https://aoir.org/reports/ethics3.pdf (Accessed December 2020).

Geller G, Boyce A, Ford DE, Sugarman J. 2010. Beyond “Compliance”: The Role of Institutional Culture in Promoting Research Integrity. Academic Medicine **85**(8):1296-1302 DOI: 10.1097/ACM.0b013e3181e5f0e5.

Harrison JR, Roberts DL, Hernandez-Castro J. 2016. Assessing the extent and nature of wildlife trade on the dark web: Wildlife Trade on the Dark Web. Conservation Biology **30**(4):900–904 DOI: 10.1111/cobi.12707.

Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, Poisot T, Woo KH, Zimmerman NB, Hollister JW. 2016. Ten Simple Rules for Digital Data Storage. PLoS computational biology **12**(10):e1005097 DOI: 10.1371/journal.pcbi.1005097.

Hinsley A, Lee TE, Harrison JR, Roberts DL. 2016. Estimating the extent and structure of trade in horticultural orchids via social media. Conservation Biology **30**:1038–1047 DOI: 10.1111/cobi.12721.

Ibbett H, Brittain S. 2019. Conservation publications and their provisions to protect research participants. Conservation Biology **34**:80-92 DOI: 10.1111/cobi.13337.

Ladle RJ, Correia RA, Do Y, Joo G-J, Malhado AC, Proulx R, Roberge J-M, Jepson P. 2016. Conservation culturomics. Frontiers in Ecology and the Environment **14**(5):269–275 DOI: 10.1002/fee.1260.

Langford D. 2000. Internet Ethics. Macmillan Education UK.

Liu JP. 2003. The DMCA and the Regulation of Scientific Research. Berkeley Technology Law Journal **18**:501–537 DOI: 10.2139/ssrn.457742.

Macfarlane B. 2009. Researching with integrity: the ethics of academic enquiry. Routledge, New York.

Markham A, Buchanan E. 2012. Ethical Decision-Making and Internet Research: Recommendations by the AoIR Ethics Working Committee. Association of Internet Researchers. Available from www.aoir.org/reports/ethics2.pdf (accessed December 2020).

Monkman GG, Kaiser M, Hyder K. 2018. The Ethics of Using Social Media in Fisheries Research. Reviews in Fisheries Science & Aquaculture **26**(2):235–242 DOI: 10.1080/23308249.2017.1389854.

National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research. 1979. The Belmont Report. Page 10. Department of Health, Education and Welfare. Available from https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf (accessed December 2020).

Newing H, Perram A. 2019. What do you know about conservation and human rights? Oryx **53**(4):595–596 DOI: 10.1017/S0030605319000917.

Nissenbaum H. 2010. Privacy in context: technology, policy, and the integrity of social life. Stanford University Press.

Nissenbaum H. 2011. A Contextual Approach to Privacy Online. Dædalus, the Journal of the American Academy of Arts & Sciences **140**(4):32–48 DOI: 10.1162/DAED_a_00113.

Office for Human Research Protections. 2020. International Compilation of Human Research Standards. U.S. Department of Health and Human Services. Available from https://www.hhs.gov/ohrp/sites/default/files/2020-international-compilation-of-human-research-standards.pdf (accessed December 2020).

Okediji RL. 2019. The Limits of International Copyright Exceptions for Developing Countries. Vanderbilt Journal of Entertainment & Technology Law **21**:689-736.

Oyama, Y et al. (Translation). 1970. Copyright Law of Japan. Available from http://www.cric.or.jp/english/clj/cl2.html (accessed December 2020).

Ricketson S. 2003. WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment. SCCR/9/7. World Intellectual Property Organization, Geneva. Available from https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=16805 (accessed December 2020).

Roulet TJ, Gill MJ, Stenger S, Gill DJ. 2017. Reconsidering the Value of Covert Research: The Role of Ambiguous Consent in Participant Observation. Organizational Research Methods **20**(3):487–517 DOI: 10.1177/1094428117698745.

Singapore Ministry of Law, Intellectual Property Office of Singapore. 2019. Singapore Copyright Review Report. Government of Singapore. Available from https://www.mlaw.gov.sg/news/press-releases/singapore-copyright-review-report-2019 (accessed December 2020).

Society for Conservation Biology. 2004, July 24. Code of Ethics. Available from https://conbio.org/about-scb/who-we-are/code-of-ethics (accessed December 2020).

Soriano-Redondo A, Bearhop S, Lock L, Votier SC, Hilton GM. 2017. Internet-based monitoring of public perception of conservation. Biological Conservation **206**:304–309 DOI: 10.1016/j.biocon.2016.11.031.

Spicker P. 2011. Ethical Covert Research. Sociology **45**(1):118–133 DOI: 10.1177/0038038510387195.

Spinello RA. 2017. Cyberethics: Morality and Law in Cyberspace, 6th edition. Jones and Bartlett Learning.

St John FAV, Brockington D, Bunnefeld N, Duffy R, Homewood K, Jones JPG, Keane AM, Milner-Gulland EJ, Nuno A, Razafimanahaka JH. 2016. Research ethics: Assuring anonymity at the individual level may not be sufficient to protect research participants from harm. Biological Conservation **196**:208–209 DOI: 10.1016/j.biocon.2016.01.025.

Sung Y-H, Fong JJ. 2018. Assessing consumer trends and illegal activity by monitoring the online wildlife trade. Biological Conservation **227**:219–225 DOI: 10.1016/j.biocon.2018.09.025.

Sutherland WJ et al. 2018. A 2018 Horizon Scan of Emerging Issues for Global Conservation and Biological Diversity. Trends in Ecology & Evolution **33**(1):47–58 DOI: 10.1016/j.tree.2017.11.006.

Tavani HT, Grodzinsky FS. 2019. Responding to Some Challenges Posed by the Re-identification of Anonymized Personal Data. Computer Ethics - Philosophical Enquiry (CEPE) Proceedings. International Society for Ethics and Communication Technology. DOI: 10.25884/jke7-mk31.

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. Biological Conservation **233**:298–315 DOI: 10.1016/j.biocon.2019.01.023.

Tulloch AIT, Possingham HP, Joseph LN, Szabo J, Martin TG. 2013. Realising the full potential of citizen science monitoring programs. Biological Conservation **165**:128–138 DOI: 10.1016/j.biocon.2013.05.025.

UK Public General Acts. 1988. Copyright, Designs and Patents Act. Available from http://www.legislation.gov.uk/ukpga/1988/48/section/29A (accessed December 2020).

Wearn OR, Freeman R, Jacoby DMP. 2019. Responsible AI for conservation. Nature Machine Intelligence **1**:72-73 DOI: 10.1038/s42256-019-0022-7.

WMA General Assembly. 2001. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Bulletin of the World Health Organization **79**(4):373–374.

World Intellectual Property Organization. 1979. Berne Convention for the Protection of Literary and Artistic Works. Available from https://www.wipo.int/treaties/en/text.jsp?file_id=283698 (accessed December 2020).

World Trade Organization. 1994. Agreement on Trade-Related Aspects of Intellectual Property Rights. Available from https://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm (accessed December 2020).

Yanco E, Nelson MP, Ramp D. 2019. Cautioning against overemphasis of normative constructs in conservation decision making. Conservation Biology **33**:1002-1013 DOI: 10.1111/cobi.13298.

Zook M et al. 2017. Ten simple rules for responsible big data research. PLOS Computational Biology **13**(3):e1005399 DOI: 10.1371/journal.pcbi.1005399.

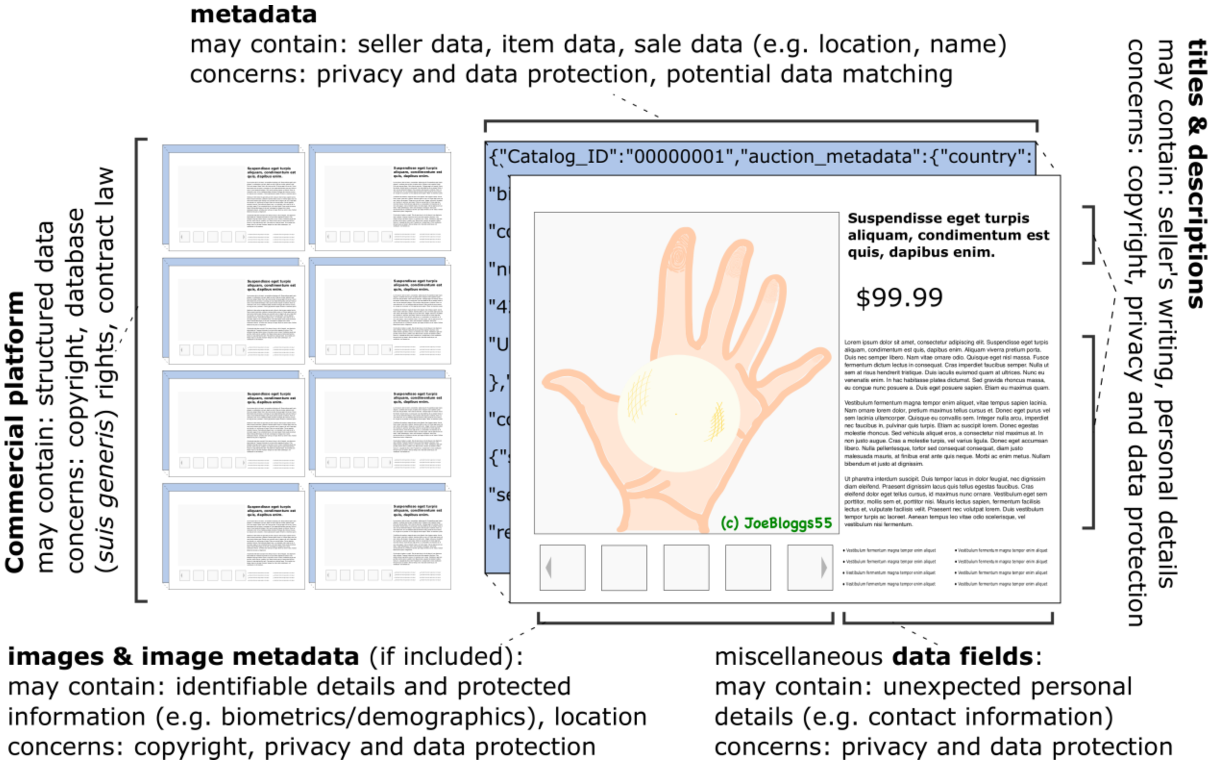

Figure 1. Online data collection requires a range of legal and ethical considerations, including privacy and data protection, intellectual property rights and contract law. Here, in the example of an auction website, automated collection methods can result in unexpected personal data collection, and published anonymized or nonpersonal data can be matched with online records for reidentification.