



FloraGuard: Tackling the Online Illegal Trade in Endangered Plants Through a Cross-Disciplinary ICT-Enabled Methodology

Anita Lavorgna¹, Stuart E. Middleton²,
Brian Pickering², and Geoff Neumann²

Abstract

This article presents a part of the ongoing Economic and Social Research Council (ESRC)-funded project “FloraGuard: Tackling the illegal trade in endangered plants” that relies on cross-disciplinary approaches to analyze online marketplaces for the illegal trade in endangered plants, and explores strategies to develop digital resources to assist law enforcement in countering and disrupting this criminal market. This contribution focuses on how the project brought together computer science, criminology, conservation science, and law enforcement expertise to create a tool for the automatic gathering of relevant online information to be used for research, intelligence, and investigative purposes. The article also discusses the ethical standards applied and proposes the concept of “artificial intelligence (AI) review” to provide a sociotechnical solution that builds trustworthiness in the AI approaches used for this type of cross-disciplinary information and communications technology (ICT)-enabled methodology.

Keywords

wildlife trafficking, plant crimes, explainable AI, ethics in online research, natural language processing

¹University of Southampton, UK

²Gamma House, Southampton, UK

Corresponding Author:

Anita Lavorgna, Associate Professor of Criminology, Department of Sociology, Social Policy & Criminology, University of Southampton, Murray Building, 58 Salisbury Road, Southampton SO17 1BJ, UK.

Email: a.lavorgna@soton.ac.uk

Introduction

Over the last 60 years, the horticultural trade, and particularly the market of exotic and wild plants, has increased significantly (Novoa et al., 2017; Sajeva et al., 2007). Exotic and wild plants are harvested and traded all over the world to use their parts and derivatives for a variety of purposes, including as pharmaceuticals, beauty products, and food. Alongside the legal trade in plants and their derivatives, the profitability of this market has contributed to the increase in illegal commerce, with endangered species traded in contravention to the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). The internet has further increased the illegal trade of plants and their derivatives, facilitating the connection of supply and demand and making it a real hybrid (online and offline) market (Lavorgna, 2014). No matter how highly specialized the market in a certain species is, it is much easier to find potential buyers or sellers online than in the physical world (Interpol, 2013, 2018; Lavorgna, 2014; Olmos-Lau & Mandujano, 2016; Sajeva et al., 2013; Wu, 2007).

Within this context, there is consensus that the policing of such criminal activity is still limited and poorly resourced (Elliott, 2012; Lavorgna, 2014; Lemieux, 2014; Runhovde, 2017). A major challenge is the fact that law enforcement agencies have limited training opportunities and lack of equipment and specific expertise to counter this illegal trade effectively (CITES, 2016; World Wildlife Fund [WWF], 2016, 2018). Crimes against wildlife have low priority on the law enforcement agenda (International Fund for Animal Welfare [IFAW], 2008); as a result, investigations are generally sparse (Fajardo del Castillo, 2016; Zimmerman, 2003). As such, there are minimal consequences for those perpetrating wildlife trafficking, making it a high-profit, low-risk criminal business (Hinsley, Nuno, et al., 2017). Consequently, the question of how we can best control and prevent this criminal market needs to be addressed.

This article presents our (United Kingdom) Economic and Social Research Council-funded project “FloraGuard: Tackling the illegal trade in endangered plants,” which analyses the criminal market in endangered plants involving the United Kingdom using mixed methods and cross-disciplinary approaches, and explores strategies to develop digital resources to assist law enforcement. For the scope of this article, we will focus on a specific part of the project, which brings together computer science, criminology, conservation science (plant ecology), and law enforcement expertise to create a tool to gather relevant online information on illegal trade in endangered plants more effectively to be used by researchers for analyses, and by law enforcement and other stakeholders for intelligence and investigative purposes. After presenting a brief outline on the illegal trades in plants and their research significance, the article continues by presenting the peculiar characteristics of the approach we are using in our (ongoing at the time of writing) research project (our sampling strategy, the data collection methods, and the strategy for data analysis). The rest of the article conceptualizes and elaborates on the information and communications technology (ICT)-enabled methodology we propose for work analyzing online markets and communities, with a specific focus on ethics. Finally, the article introduces the idea of an “Artificial Intelligence (AI) review” to be included in sociotechnical studies relying on AI

techniques to facilitate critical awareness of the potential for bias in the use of AI algorithms.

Background

Plant crimes have long been a focus of concern mainly in conservation science (among others, Dobson, 1996; Goetsch et al., 2015; Hinsley, Nuno, et al., 2017; Hinsley & Roberts, 2018; Phelps & Webb, 2015; Regan, 2004; Sajeve & Carimi, 1994; Sajeve et al., 2013). In criminology and other relevant disciplines, while the illegal trade in wild animals (and animal parts, such as ivory) has been under the spotlight especially over the last decade (see, among others, Felbab-Brown, 2017; Moreto & Lemieux, 2015; Sollund, 2015; van Uhm & Wong, 2018; Wyatt, 2013), the illegal trade in plants has so far been relatively underinvestigated (with some notable exceptions, for example, Arroyo-Quiroz & Wyatt, 2019). Indeed, most discussions about crimes against wildlife (in academia, but also in policy making and consequently on law enforcement policies) remain limited to reducing the trade of charismatic endangered animals and their derivatives (such as ivory from elephants) (IFAW, 2017). On the contrary, the global illegal trade in plants continues to receive little attention (Margulies et al., 2018, 2019)—an issue that has been described as “plant blindness” (Balding & Williams, 2016; Wandersee & Schussler, 1999). The main exception to this blindness is timber, the illegal trade of which is receiving increasing attention; this can be explained by the superior monetary value of the timber trade, as well as by the visible impact of logging on forested ecosystems (Margulies et al., 2019). The importance of reducing illegal plant trades, however, is not a problem that should be overlooked: it has been recognized that wild plant trafficking threatens and destroys numerous species, causing important conservation problems (Head et al., 2014; Herbig & Joubert, 2006; Phelps & Webb, 2015; Schippmann et al., 2002), and hinders the rule of law, security, and good governance (Haenlein & Smith, 2017; The United Nations Office on Drugs and Crime [UNODC], 2016; WWF, 2016).

As noted in “Introduction,” the main legal framework regulating illegal plant trade is the 1975 CITES, which aims to control the trade of species in which international commercialisation and over-exploitation poses a serious threat to their survival, or could threaten survival in the future if not regulated (Lavorogna et al., 2018; Young, 2003). CITES provides three levels of protection for endangered species, depending on the level of threat: species listed in Appendix I are threatened with extinction and the trade in wild specimens is permitted only in exceptional circumstances; species included in Appendix II are considered vulnerable; therefore, trade in wild collected specimens is allowed, but is subject to the issuance of a permit; species listed in Appendix III are protected in at least one country, which has asked other CITES parties for assistance in controlling the trade; as such, trade in wild collected specimens is permitted, but is subject to export permits or certificates of origin. To assess how much a plant is endangered, the most authoritative reference point is the International Union for Conservation of Nature (IUCN) Red List of Threatened Species, a science-based inventory of the global conservation status of plant and animal species, which is based

on a set of quantitative criteria to evaluate the extinction risk of thousands of species (Brummitt et al., 2015; IUCN, 2018).

More than 30,000 taxa of plants are included in the CITES Appendices (vs. only about 5,000 taxa of animals). Overall, CITES has encouraged the artificial propagation of many plants, reducing the pressure on wild populations (Sajeve et al., 2007). However, some plant collectors as well as some buyers in the medicinal plant market still prefer wild-collected plants because there is a kind of thrill or status to owning a rare plant from the wild that others cannot have (vs. mass produced nursery plants), or these plants are believed to contain superior active ingredients (McGough et al., 2004; Royal Botanic Gardens, Kew [RBG Kew], 2017; Sajeve et al., 2013).

Despite the slight increase in studies on the illegal trade in plants (and especially on endangered cacti and orchids) in recent years (see among others Arroyo-Quiroz & Wyatt, 2017, 2019; Hinsley, de Boer, et al., 2017; Hinsley Nuno, et al., 2017; Olmos-Lau & Mandujano, 2016), there are still significant knowledge gaps that need to be dealt with for better control and to prevent the problem. One of these gaps is the role of the internet as a facilitator in the illegal trade of plants and their derivatives; a better understanding of the characteristics of online illegal markets and of the actors operating in them is in fact a necessary starting point for any intervention to mitigate the problem, and for the initiation of any effective investigation. Only a limited number of studies have looked specifically at online criminal markets. These studies have invariably focused on data from secondary sources (investigative transcripts, judicial material, media and gray literature) (Lavorgna, 2014), or collected data from specific online markets (such as a limited number of generalist auction sites, see CITES, 2017; Wu, 2007) or on specific taxa (Harrison et al., 2016; Hinsley et al., 2016; Krigas et al., 2017). Furthermore, these studies have not given detailed attention to studying the socioeconomic and sociobehavioral characteristics of markets and sellers respectively, which is particularly important in exploring the possibilities for online crime prevention and effective online policing (Brenner & Clarke, 2005; Grabosky & Smith, 2001). Also, they have not focused specifically on policing aspects in tackling the illegal trade in endangered plants, even though the lack of proper datasets for wildlife trafficking is regrettable since it is undermining supply-chain monitoring and fast aggregation of data to inform policy-making (Chan et al., 2015). It is well recognized that online trades should become the priority for law enforcement improvement (Hinsley, Nuno, et al., 2017; Kretser et al., 2015), and interventions for capacity-building of enforcement personnel to increase the quality of market surveillance for illegally traded wildlife. Further the creation of easy-to-use tools and resources to aid in the identification of traded wildlife and products has been called for (Kretser et al., 2015).

In our study, we are focusing on the United Kingdom. There is consensus that most wildlife is trafficked from developing countries to the Western world, with Europe (and the United Kingdom) considered one of the top global importers by value of wildlife, and as a transit hub for trafficking to other regions (Engler & Parry-Jones, 2007; European Union [EU] Commission, 2016a; UNODC, 2016). However, there is a general lack of reporting and prosecutions especially for plant crime. In the United Kingdom, but most likely in many other countries as well, this is due to the inability

to produce information on how prevalent offending is, and where hotspots may be. Consequently, very limited actual data are available to analyze to identify criminal trends (Link, 2016).

Multi- and cross-disciplinary approaches to the prevention, detection, and control of wildlife trafficking (but not focusing specifically on plants) are yielding promising results, as they bring together diverse expertise to address the complex challenges associated with the preservation of natural resources (Alacs et al., 2010; Cowan et al., 2006; Gibbs et al., 2010; Haas & Ferreira, 2015; Lemieux, 2015; Staats et al., 2016). When it comes to the detection of illegal wildlife trade online, there have been some recent multi- and cross-disciplinary efforts bringing together conservation science and sustainability science expertise with web/computer science or social science methodological approaches. Di Minin and colleagues (2018), for instance, suggest that automated systems based on machine learning approaches, such as deep learning, neural networks, and natural language processing, can be used to detect illegal commodities on a large-scale, as they can identify verbal, visual, and audio-visual content pertaining to illegal wildlife trade. However, the authors recognize that effective results cannot be achieved without liaising with experts in particular taxonomic groups. In looking at how NGOs, government agencies and academics are linked together to tackle illegal wildlife trade (including of plants), researchers at the University of Kent (Moshier et al., 2019) have carried out a social network analysis using data gathered from online questionnaires (72 institutions) and semistructured interviews (11 individuals). This work was conservation-science led and used statistical and commercial network analysis tools (e.g., UCINET 6)¹ to explore stakeholder communities. Supervised machine learning was used in work by Carl Miller and colleagues (Miller et al., 2019) to identify online marketplaces and websites involved in the trade of a selection of CITES-listed animals and plants. They used the *Bing* search engine to identify 121,741 web pages from 3,329 sites and extract key phrases from these pages related to themes such as sales-related conversations, mentions of known sellers, and purchasing-related discussion phrases. These results were then used to train supervised machine learning classification algorithms to visualize the statistical breakdown of online activity by class (e.g., number of vendors mentioning orchids).

Our Approach

In FloraGuard we explore how natural language processing techniques—such as open information extraction (Banko et al., 2007)—could be used to gather information from large crawls (i.e., a structured set of data obtained by a software program visiting a web page and grasping content and links from it) of auction and forum websites. We aim to analyze the information crawled to identify patterns at a (online) community level, able to plot (i.e., representing the data through a graph) the extent and evolution of trades in target plants (see below for the sampling strategy used), while preserving links to original evidence for follow-up via police investigation.

The approach to crawling is motivated by Miller and colleagues (2019), but our use of open information extraction algorithms to extract tuples, each with grammatical subject, predicate and object phrases, from posted text allows a much finer grained linguistic analysis of online posts based on the immediate syntactic context than has been previously possible in systems limited to keyword extraction. The FloraGuard computer science tools integrate into a wider sociotechnical system, underpinned by an ICT-enabled methodology which brings together criminology, computer science, and conservation science tools and methods in a rich cross-disciplinary working environment.

Sample Selection

Our first task was to select a sample of online marketplaces for data gathering and analysis based on a maximum variation sampling strategy for wild plant trade behavior types. “Data” in this context are all types of information present in these marketplaces, such as information on products, forum members, trade-related services offered, links to external websites, and the language/jargon/argot used.

To select our sample, we decided that relying only on the marketplaces identified by the existing literature (e.g., eBay) was not enough: first, market dynamics online can change quickly, as illegal activities can move with relatively little cost and effort from one platform to another; second, looking solely at known marketplaces would have limited the aim of our approach, which instead hopes to uncover new marketplaces that have otherwise been overlooked by both researchers and practitioners. Furthermore, beyond looking at generalist marketplaces (those, such as eBay, selling a broad variety of products including but not limited to several types of plants and/or their derivatives) we wanted to access specialized marketplaces as well. Indeed, many types of plants and their derivatives are of interest to very specific groups of customers because they are enthusiastic about a certain species (for instance, orchid lovers) or because of the characteristics of a certain derivative product (for instance, slimming pills). Consequently, they can be marketed exclusively to a very specific segment of the population through specialized forums.

Hence, to identify our sample of online marketplaces, we decided to focus on a limited number of selected genera and species, which were deemed to be particularly relevant for our study as they met the following criteria: (a) they are a priority from a conservation science point of view since threatened with extinction, and therefore they are listed in the CITES Appendix I² and/or considered endangered or close to threatened by the IUCN; (b) the genera and species selected include sufficient variety of traded “products,” meaning that we aimed to take into consideration both plants traded as live specimen and as derivatives, and both for horticultural and for health-related purposes; and (c) there was sufficient volume of online posts available so we could examine how quantitative results from large web crawls could best provide supporting evidence for criminological hypotheses. In this way, we expected to find marketplaces with different characteristics and to be able to explore the strengths and weaknesses of our proposed ICT-enabled methodology across the variations within our sample set.

The criteria used, as well as the genera and species chosen, were identified by an initial preassessment carried out by a research assistant with expertise in plant ecology, based on an examination of relevant literature and exploratory online searches for specific products and species. The assessment was then refined through consultations with our project partners—namely the Royal Botanic Garden Kew and the UK Border Force CITES Team—consultations with our nonacademic project advisors,³ and through a total of 15 interviews with team leaders and senior law enforcement officers in customs, cybercrime, and wildlife crime units in the United Kingdom, as well as with relevant experts from national and international NGOs and other institutions working on wildlife trafficking. FloraGuard researchers in criminology and computer science were trained by the Royal Botanic Garden Kew experts on the species selected, and to identify proxy indicators of illegality of specific specimens (e.g., by looking at their dimension, color, or root characteristics) so as to be able to recognize signs of illegality in online trades better (e.g., when photos are posted).

The resulting selection includes:

- *Saussurea costus* (species)—from India and Pakistan, mostly collected in the wild because of the cultivation cycle and high-volume trade, traded mainly in derivatives for medical use though live plants might also be traded occasionally.
- Eight species (and subspecies) of *Ariocarpus* (genus) that are listed in CITES Appendix I and that have an IUCN Red List Status—from north Mexican highlands and Texas, traded mainly as live plants (they are relatively easy to ship over long distances) but also in derivatives for medical use.
- Eight species of *Euphorbia* (genus) that are listed in CITES Appendix I and that are considered critically endangered or endangered by the IUCN—from Madagascar, traded mainly as live plants and seeds for horticultural trade.
- A small selection of six species of cycads (genus) that are listed in CITES Appendix I and that are considered vulnerable or close to threatened by the IUCN—from the African continent and especially South Africa, traded mainly in derivatives for medical use.

Data Collection

Data were collected both from forums and from marketplaces. For forums, first of all a search was carried out using the *Bing* Search API to find a list of possible sites. *Bing* was used instead of *Google* as the *Google* API at the time of writing imposes a limit of 100 results per query and does not permit more than 100 queries per day. A *Bing* Search was carried out for plants being studied, grouped by genus/species, using a collection of search terms reflecting the vocabulary used in plant trades. A list of relevant keywords encompassing other official names in CITES listings as well as common names and vernacular names was provided by experts in plant ecology at Kew. Combinations of keywords to be used and identifications of words to be blacklisted to make the online searches more effective were manually tried

Table 1. Search Terms for *Ariocarpus* Group.

Names	Excluded terms	Forum terms	Buy/sell terms
<i>Ariocarpus</i>	seed(s)	Forum	Web + buy
<i>Ariocarpus</i> + <i>agavoides</i>		thread	internet + buy
<i>Ariocarpus</i> + <i>bravoanus</i>		Topic	Buy
<i>Ariocarpus</i> + <i>hintonii</i>			Buy + online
<i>Ariocarpus</i> + <i>kotschoubeyanus</i>			Order
<i>Ariocarpus</i> + <i>kotschubeyanus</i>			Sale
<i>Ariocarpus</i> + <i>kotschobeyanus</i>			Selling
<i>Ariocarpus</i> + <i>kotsch</i>			Purchase
<i>Ariocarpus</i> + <i>confuses</i>			Swap
<i>Ariocarpus</i> + <i>albiflorus</i>			Barter
<i>Ariocarpus</i> + <i>retusus</i>			
<i>Ariocarpus</i> + <i>scapharostroides</i>			
<i>Ariocarpus</i> + <i>scapharostrus</i>			
<i>Ariocarpus</i> + <i>scaphitostris</i>			
<i>Ariocarpus</i> + <i>trigonus</i>			
“Tamaulipas living-rock”			
“Nuevo leon living-rock”			

independently by two researchers (one in criminology and one in computer science), who then compared and refined them. For instance, in some cases, specific terms (such as seeds) were excluded since we wanted to target live plants or derived products from live plants that were likely to be traded illegally—a decision taken on a case-by-case basis, after discussion with conservation science experts. Additional sets of search terms were used to search for forums/marketplaces and discussions regarding buying and selling after exploratory manual online investigations. In the searches we included common misspellings, although modern search engines have some tolerance for misspellings already.

The search terms used for the *Ariocarpus* group are shown in Table 1. A “+” indicates that all terms are required (in any order), and quotation marks indicate that an exact sequence is required. Mostly exact sequences are not required as users frequently use terms such as “*A. agavoides*.” It is only possible to tell that they are referring to *Ariocarpus* if the word “*Ariocarpus*” appears elsewhere in the document.

It is worth noting that, while our searches are targeted to identify English-speaking actors and communities, there is some evidence (from research carried out by other researchers on eBay) that the code words and phrases associated with online wildlife trades are often consistent across various countries and languages, suggesting that the globalization of commercial hotspots may be homogenizing conventions within trading communities (Alfino & Roberts, 2018). This consistency may offer new opportunities for tackling wildlife trades online, but further multilingual research on different marketplaces and forums is necessary to corroborate this finding.

Table 2. Data Characteristics Found for *Ariocarpus* on a Large Forum Site.

Tot number threads	Tot number pages	Tot number posts	Tot number unique users	Tot number words	Mean word count/post
646	2,511	32,493	1,562	1,505,122	46

Through the consultations and interviews mentioned above we discussed the main or emerging marketplaces as identified, for instance, by recent or ongoing investigations, to make sure that our searches were not missing major or otherwise important known marketplaces. Furthermore, after in-depth discussion with interviewees and project partners we decided to limit our analysis to the clear web because of lack of any evidence of significant plant trafficking in parts of the web not reachable by normal search engines (dark web).⁴ This is in line with the findings of Harrison and colleagues (2016) and Roberts & Hernandez-Castro (2017), who have shown that illegal wildlife trade takes place over the dark web only in very small quantities, generally as bycatch (i.e., when the products are potentially illegal for other reasons, such as cacti traded for their hallucinogenic properties). This is probably not surprising: as enforcement in the clear web is very limited, there is very little incentive for traders to move onto the dark web, where their pool of potential customers might be more limited (Lavorigna, 2014; Roberts & Hernandez-Castro, 2017).

Once a set of forum site homepages has been found through *Bing* search, the most relevant ones are selected for crawling using Undercrawler,⁵ a Scrapy-based web crawler developed by Defense Advanced Research Projects Agency (DARPA). Undercrawler has the benefit of being able to perform actions such as automatically logging into websites, carrying out searches and handling pagination—tasks that are difficult for traditional web crawlers. Each forum homepage is crawled for discussion threads, then pages from discussion threads containing relevant keywords are downloaded and parsed (based on HTML paths and regex) to extract both the text of each post and relevant metadata. Forum user profile pages are also parsed if they are active in crawled discussions. Where forum Terms and Conditions (T&C) permit crawling, we create forum accounts for Undercrawler to use, otherwise we rely on the public pages *Bing* search finds alone.

For example, in the *Ariocarpus*-related searches we identified five main forums where we found over 400 relevant threads. Table 2 shows the breakdown of crawled posts from one of these larger forums. Depending on the popularity of species and forum site, there can be a lot fewer threads (e.g., under 100) to crawl.

For marketplaces we focussed on four big marketplaces known to be used currently in the illegal plant trade following discussions with project partners and interviewees, and exploratory manual online investigations: Ebay, Amazon, Etsy, and Alibaba. For example, on Etsy the *Ariocarpus* group search found 13,815 products, while the *Saussurea* group search found 3,430 products. On manual inspection of a random sample of posts we found the majority were not about live plants, and some posts were about other types of cacti and succulent genera, so it is clear that further filtering will

be required to reduce false positives. This work is ongoing; unfortunately, in the context of this research it would be unrealistic to eliminate false positives without incurring in the opposite (and worst, for the scope of our work) problem of false negatives. However, we believe that thanks to the cross-disciplinary ICT-enabled methodology explained in the dedicated section below we have limited this issue in a satisfactory way.

It should also be noted that not all online conversations can be crawled. In some cases, a plant for sale may be introduced in the forum publicly, but the actual sale may be conducted through private messaging between members (Franklin et al., 2007; Holt et al., 2016). On marketplaces, most communication between the seller and the buyer is private with feedback visible, but partially anonymous. On Alibaba, for example, only the first and final letter of the usernames of feedback providers are given.

Data Analysis

At the time of writing, we have crawled data for three plant groups. For instance, for *Ariocarpus*-related searches, as anticipated above we crawled a total of five major forums (for a total of 52,217 posts), three of which focusing on horticulture and two focusing instead on drug-related and ethnogenetic aspects of the trade (entheogens is the use of fauna, flora, and fungi species with psychoactive properties in cultural, religious, shamanic, or spiritual contexts). Indeed, probably because *Lophophora Williamsii* (or peyote, which is sought after for its hallucinogenic effects) is a cactus species that occupies a similar range spanning both Mexico and Texas, we discovered that also some *Ariocarpus* are sought after and traded for their presumed psychotropic properties.

We are currently looking for observable and repeatable patterns exhibited online around the illegal plant trade in these forums that we could use to train artificial intelligence tools, as well as to improve our understanding of the criminal market under investigation. Observed behaviors and attributes associated with online trade to be observed include selling mechanisms (e.g., buy-it-now, auction sites, forums) and types (one-off special trade vs. the sale of bulk trade items); actor roles (e.g., vendor, customer, community product expert providing feedback, and advice); payment mechanisms (e.g., PayPal; bank transfers; cash-in-hand; use of private messaging for trade completion) and types (fixed price and price ranges for product); location of the product (country of origin or of trade); product exchange location (e.g., shows and events, through the postal service); shipping information (worldwide, European Union only, United Kingdom only); mention of permits (CITES mentioned; phytosanitary permit mentioned; no permit mentioned; *caveat emptor* mention); social interaction types (e.g., advert; expression of interest; reassurance prior to trade; feedback on trade; barter on price/make an offer; discussion of trade location; exchange of private messaging details; share technical information about avoiding police/customs). Other observed aspects of sociological relevance include: subcultural features; framing of conservation-related matters; framing of CITES-related matters; techniques of neutralization; drug use narratives. Throughout this process we are using a cross-disciplinary

approach, with criminological expertise used to review the initial AI robot data patterns, and to corroborate them by performing manual investigations and qualitative data analyses on purposely selected data samples.

Developing a Cross-Disciplinary ICT-Enabled Methodology

Traditional criminological analysis of online communities (e.g., Holt, 2007; Holt & Lampke, 2010; Lavorgna & Di Ronco, 2017; Mann & Sutton, 1998; Steinmetz, 2016) has involved a mixture of sampling of online posts and analysis of questionnaire and/or interview data. Sampling is typically performed on manually downloaded online webpages and/or transcripts from interviews of key stakeholders (Holt, 2017). Sampling strategies (ranging from random sampling and maximum variance sampling to convenience sampling) are carefully assigned to each task. A set of case studies are often developed, and then evidence mined and annotated from the sample set and linked to case study themes supporting hypotheses under development. Tools such as NVivo or NodeXL are used to curate and annotate document sets or to perform statistical and network analysis of people, locations, and behavioral types. Evidence in the form of statistical patterns is also used to support or weaken hypotheses, especially around causation analysis, often leading eventually to the development of recommendations for policy change such as proposals for new online intervention strategies.

Previous criminology-led work “borrowing” computer science approaches to crawl and analyze cybercrime posts have focussed on keyword searches to find relevant pages, and relied on a combination of heuristics and forum category metadata to identify entities being discussed. When analyzing online marketplaces such as drug markets, recent approaches (Ball et al., 2019; Décary-Héту & Aldridge, 2015; Demant et al., 2018) have used crawlers using a combination of forum HTML metadata and/or regex patterns to identify drug names and financial information. To increase data quality manual analysis of information from small sample sets of posts is also used (e.g., top N posts for best-selling drug products). Data analysis tools such as the Jupyter Notebook were used to provide subsequent statistical analysis.

Access to AI algorithms from the computer science domain, including machine learning and natural language processing, allows us to explore a wider sociotechnical system for such work. In FloraGuard, we wanted to examine how traditional criminology methodologies could be “ICT-enabled,” augmenting the capabilities of the criminologist with an ability to gather quantitative evidence at scale from online communities. Such sociotechnical systems are by their nature cross-disciplinary. Figure 1 shows at a conceptual level our proposed ICT-enabled methodology for work analyzing online communities.

Intelligent bootstrapping and distant supervision (Smirnova & Cudré-Mauroux, 2018) and transfer learning (Pouyanfar et al., 2018) across different domains are key challenges for computer science areas, and appear in subdomains such as data mining, machine learning, and computational linguistics research. Unsupervised bootstrapping

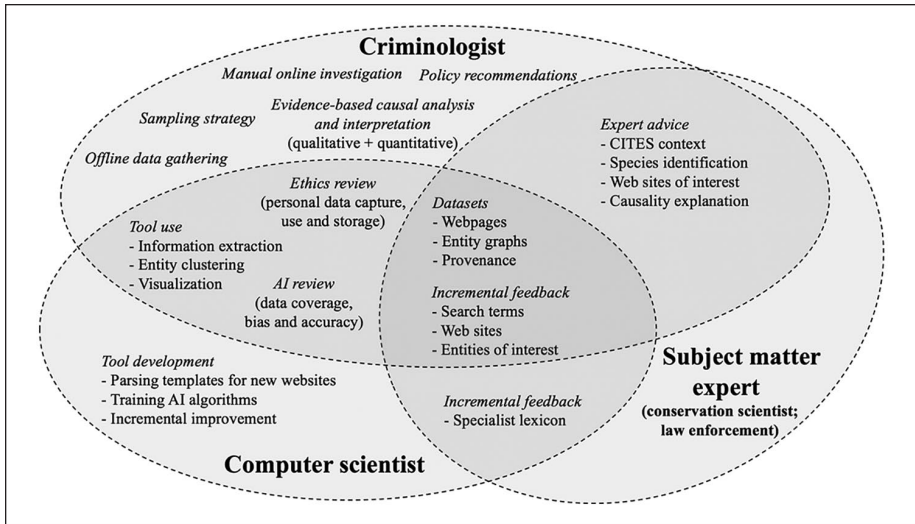


Figure 1. ICT-enabled methodology in FloraGuard.
 Note. ICT = information and communications technology.

algorithms generate seed annotations from an unlabeled training set, creating a labeled training set that allows subsequent application of higher precision semisupervised algorithms. Distant supervision uses an existing knowledge-base, such as Wikipedia knowledge graphs as a source of training data. Transfer learning uses approaches such as inductive transfer, precomputer word embeddings such as Bidirectional Encoder Representations From Transformers (BERT) (Devlin et al., 2018), attention transfer or representational learning to apply patterns learnt on one domain to data sets in a similar domain. The bootstrapped open information extraction algorithms being developed within FloraGuard have been designed to work within an ICT-enabled methodology and not work as a black box. By using features originating from grammatically coherent propositional structures (e.g., subject: “John”; predicate: [“wants to buy,” “*Ariocarpus*”]) to extract webpages entities of interest (namely people, locations, buying behaviors, and species), we are able to provide some explanation for our quantitative patterns and provide a direct provenance trail from features back to the original online posts which generated them. This means that we can trace back the connections between the entities extracted and the original webpages, something that can be of great importance, for instance, for investigative purposes. This type of feature provenance trail is not preserved in many of today’s information extraction approaches, especially those based on recurrent neural network (RNN) where explainability of black box AI solutions is still an unsolved active research challenge.

Through the FloraGuard project, we are exploring how our tools can make use of incremental feedback during the development of criminological data exploration and hypothesis development and testing. As will be explained in more detail in the following

section, initially subject matter experts (conservation scientists/plant ecologists and law enforcement, in our case) provided lexicons relating to specialist vocabulary for the domain, which is important to augment large generic lexicons often sourced from previous projects crawling the web. Some exploratory manual online investigations were carried out by the criminologist to suggest additional market-related vocabulary and potential words to be used for search or filtering. In this way, computer scientists were able to access in-depth domain expertise from criminology and conservation science researchers relating, for instance, to the plant/illegal online trade specialist vocabulary (e.g., plant types and values, marketplace social norms, typical vendor interaction patterns). The crawler search terms evolve as the criminological analysis evolves, with new entities of interest appearing such as people (e.g., vendors of interest), species, locations, and behavior markers (e.g., keywords associated with a trade type, such as “sale” or “barter”).

AI Review

As data sets will be populated incrementally as our analysis evolves, we understand they will be sparse and fragmented initially at least. This means AI tools working with this data will inherently contain data bias. We wanted our methodology to embrace this aspect and promote AI trustworthiness. The “incremental feedback” system described in Figure 1 allows us to work in this direction, as it allows a constant interplay between criminology, computer science, and subject-matter expertise, and encourages, via qualitative insight and specialized human knowledge, the mitigation of any excessive generalization or streamlining introduced by the automation of the data gathering and analysis processes.

In addition to this cross-disciplinary effort, and to using explainable and provenance-preserving AI techniques, we propose that in similar studies an “AI review” step should be included to review the wider sociotechnical system. This is something analogous to the ethics review regularly conducted for studies involving human subjects and personal data (see dedicated section below) and would allow a panel of experts to periodically review potential bias in sampling and training data that AI systems are using. This review would allow us to look at AI algorithms and their potential for bias especially in the context of incrementally crawled data where outputs will always start off as sparsely populated and fragmented until the intelligence picture develops and the right search keywords are discovered to create good online forum coverage. Key to this is finding ways to keep provenance links from results back to original source posts, and making the final decision makers, who will receive results as an intelligence package, aware of any limitations in coverage behind online crawled data. The aim is to ensure adequate measures are taken such that stakeholders making decisions based on results from the AI system are aware of any bias and have taken appropriate mitigation strategies (e.g., sourcing additional corroborating human evidence for patterns based on data segments known to be demographically underrepresented). “Datification” is known to be potentially problematic in estimating crime patterns and guiding policy change (Chan & Bennet Moses, 2016; Ferguson, 2012;

Lavorgna, 2020; Shapiro, 2017; Williams et al., 2017); as computational criminology and cross-disciplinary approaches are becoming increasingly common, issues related to datification are bound to increase as well. As such, the proposed concept of “AI review” hopes to provide a sociotechnical solution that builds trustworthiness into the AI approaches it uses.

Ethical Considerations in Data Collection

Research Ethics Committees (REC, or Institutional Review Board, IRB, in the United States) have the responsibility to advise, review and monitor ethical research practice. Professional bodies similarly issue guidelines to support their membership both in professional practice and research. It has often been assumed that traditional processes are suitable for direct translation and application to the virtual world (Beaulieu & Estalella, 2012; Coudert, 2010). In consequence, the concepts of informed consent, expectations of privacy and anonymity would be assumed to hold. At the same time, the gradual realization that mediated, and especially internet-based, research does differ has led to renewed debate and some tentative proposals (Ess, 2002; Markham & Buchanan, 2012). Recent cases of covert social-network-based experimentation (Flick, 2016) and the much-publicized Cambridge Analytica case (Cadwalladr & Graham-Harrison, 2018; Ward, 2018) have raised significant ethical questions not only among academics, but also the prosumers of online content themselves.

The moral imperative to control the illegal trade in plant life given our collective responsibility for the environment seems clear-cut. Combining techniques from data analytics with more established study of online fora and chatrooms provides criminologists with significant advantages in understanding the exploitation of internet resources and the extent of the problem with illegal trade. However, for the RECs reviewing and providing institutional approval for this research, the project requires the continued engagement with the broader debate about the ethics of online research.

The Belmont Report,⁶ the New Brunswick Declaration (Van Den Hoonaard, 2013), the European Code of Conduct for Research Integrity (ALLEA, 2017) and the Declaration of Helsinki (Carlson et al., 2004) all stress fundamental ethical principles of respect for the individual (the participant, as well as subsequently the researcher) balanced by the potential benefit to the community of the research. Online, this has been translated to the following concepts:

- Do the participants expect the virtual space (e.g., a forum) to be *public* or *private*? Some will not engage online in possibly illegal activity for fear of being caught (Barratt et al., 2013). Some expect a safe place to be who they want without constraint (Suler, 2004).
- Should the researcher seek *informed consent* from the participant(s)? This may not be practical or without risk to the researcher (Décary-Héту & Aldridge, 2015), or seen as intrusive (Sugiura et al., 2017).
- How can the researcher protect the *anonymity* of the participant(s)? This may not be appropriate for perfectly valid and prosocial reasons (Michael et al.,

2006) or practically impossible given the power of search engines and persistence of content (Beaulieu & Estalella, 2012; Sugiura et al., 2017).

At the same time, public and NGO concerns are growing in the face of algorithmocracy and datification. The EU General Data Protection Regulation (GDPR), for instance, sought to give control back to the data subject to restrict processing and query automated decision-making and profiling (EU Commission, 2016b). Nonetheless, there is now increasing concern about the potential for mis-categorization as the result of those opaque algorithms (Cheney-Lippold, 2017; O'Neil, 2016).

So, the RECs (IRBs) now look to the community of practice for guidance. Requests for approval are currently reviewed on a case-by-case basis. Markham and Buchanan (2012) and the AoIR suggest that the research must be contextualized rather than simply consider *private versus public spaces*, *informed consent*, and *anonymity*. Like Michael and colleagues (2006) had already proposed, within a general framework, there needs to be some regard for the benefit both to participant and to the community. Most importantly, there needs to be a continued dialogue with the community at large to establish and re-visit what is regarded as acceptable practice (Coudert, 2010; Markham and Buchanan, 2012; Van Den Howard, 2013).

To contextualize, for obtaining RECs' (IRBs') approval it is important to consider the T&C of the source site or sites, and the nature of engagement on those sites (is it subscription-based, or free-to-access? And so forth). This may also inform consent: if users are posting personally sensitive content, then what safeguards are in place? Would a request for informed consent disrupt the natural flow of information on the site? And would it lead to increased risk for the researcher? Finally, for illegal activity, we need to consider the social implication of leaving such activity unchecked. With continued discussion and the exchange of our research experience with our peer institutions, RECs (IRBs) and researchers are gradually moving toward a common understanding of the implications for exploiting the potential of online activity.

In the context of our ICT-enabled methodology we submitted our data study plans to our university REC for approval (approved—ERGO/FPSE/41,260 and ERGO/FPSE/46393). For FloraGuard, we do not engage in participant observation of virtual communities, but rather in their passive monitoring and in downloading of data created by online community users. Consequently, we are not engaging in any entrapment activity or encouraging the illegal trade; we have never tried to identify specific people performing specific illegal acts, nor to purchase any illegal plants, nor physically to meet anyone. In this way, we avoid risks related to the deception of participants in virtual communities or interference with unknown law enforcement operations taking place in the selected communities (Holt & Lampke, 2010), as indicated by existing research on online crime markets (Décary-Héту & Aldridge, 2015). The strict policy of passive crawling is also needed to protect staff from all potential risks relating to remote engagement with potential offenders and to reduce contamination bias. Hence, manual browsing was performed without any posting or interactions (e.g., liking a post or rating a user), and data was automatically crawled using read-only requests. Where we need to login to forums, and therefore to agree to T&C around crawling data, we have

honored any robot crawling policy. For sites that do not allow crawling in their T&C, we have only used public pages discovered by search engines without using a forum login (which, interestingly, is in line with what law enforcement working on illegal wildlife trades can currently do, as they do not have legal permission to consider in an intelligence package what is not available to them through the open webpages unless they obtain separate authority such as a search warrant).

Conclusion

From the outset, the FloraGuard project aimed to respond to the call of environmental and antiwildlife crime agencies to explore alternative and cross-disciplinary options for the prevention and control of the illegal trade in endangered species (EU Commission, 2016a; UNODC, 2016; WWF, 2016, 2018) by bringing together expertise from criminology, computer science, conservation science, and law enforcement. We hope that the outcomes of the methodological approach presented in this article can be of practical use to the work of law enforcement (such as national wildlife crime units and customs officers) and other relevant stakeholders in identifying cases of illegal online trade in prohibited (and especially endangered) plants and their derivatives, as well as in fostering the more general improvement of awareness and technical capacity in investigation and prosecution services for wildlife crimes.

From a more academic perspective, by integrating insights and expertise from criminology, computer science and conservation science, the methodology presented has important implications for demonstrating cross-disciplinary methodological developments. Furthermore, the proposed ongoing iteration of qualitative expertise from social scientists evaluating the quantitative output from data and computer scientists will, we believe, ensure appropriate conclusions are drawn from the data. There is consensus that cyberspace is increasingly important to access and monitor traces of meaningful activities and behavioral patterns, given the right tools and techniques. The approach presented in this paper could be adapted to the study of other online (criminal) activities, and we hope that the considerations around ethics and the suggested AI reviews might be useful to other researchers using similar cross-disciplinary ICT-enabled methodologies.

Acknowledgments

We thank our research assistant Catherine Rutherford for her precious work in the preliminary stages of this research. We also thank Valentina Vaglica and Carly Cowell from the Royal Botanic Garden Kew for their continuous support, and Carly also for her comments on an earlier draft of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Economic and Social Research Council [ES/R003254/1].

Notes

1. See <http://www.analytictech.com/archive/ucinet.htm>.
2. In this way, we could expect to identify specimen/products to be traded illegally, as the commercial trade in wild-collected specimens of species listed in Appendix I (that is, species threatened with extinction) is generally prohibited. It is permitted only in exceptional circumstances, for example, import/export permits issued if the specimen is not to be used primarily for commercial purposes and that such import/export is not detrimental to the survival of the species.
3. See <http://floraguard.org/people/> for further details.
4. On the contrary, there is consensus that there are other “hidden” areas of cyberspace (such as Facebook closed or secret groups) where illegal wildlife trafficking takes place. Unfortunately, as explained more in detail in the section on ethics later on in this article, our research team did not have ethical approval to search and analyze them, and also law enforcement officers have major limitations in accessing them.
5. See <https://github.com/TeamHG-Memex/undercrawler>.
6. See https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf.

Reference list

- Alacs, E. A., Georges, A., FitzSimmons, N. N., & Robertson, J. (2010). DNA detective: A review of molecular approaches to wildlife forensics. *Forensic Science, Medicine and Pathology*, 6, 180–194.
- Alfino, S., & Roberts, D. L. (2018). Code word usage in the online ivory trade across four European Union member States. *Oryx*. Advance online publication. <https://doi.org/10.1017/S0030605318000406>
- ALLEA. (2017). *2017: The European code of conduct for research integrity*.
- Arroyo-Quiroz, I., & Wyatt, T. (2017). *Green crime in Mexico: A collection of case studies*. Palgrave.
- Arroyo-Quiroz, I., & Wyatt, T. (2019). The illegal wildlife trade between Mexico and the European Union. *International Journal of Crime, Justice and Social Democracy*. Advance online publication.
- Balding, M., & Williams, K. J. (2016). Plant blindness and the implications for plant conservation. *Conservation Biology*, 30(6), 1192–1199.
- Ball, M., Broadhurst, R., Niven, A., & Trivedi, H. (2019). *Data capture and analysis of darknet markets*. <http://dx.doi.org/10.2139/ssrn.3344936>
- Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007). Open information extraction from the web. In R. Sangal, H. Mehta, & R. K. Bagga (Eds.), *Proceedings of the 20th international joint conference on artificial intelligence (IJCAI'07)* (pp. 2670–2676). Morgan Kaufmann Publishers.
- Barratt, M. J., Ferris, J. A., & Winstock, A. R. (2013). Use of Silk Road, the online drug marketplace, in the United Kingdom, Australia and the United States. *Addiction*, 109(5), 774–783.

- Beaulieu, A., & Estalella, A. (2012). Rethinking research ethics for mediated settings. *Information, Communication & Society*, 15(1), 23–42.
- Brenner, S. W., & Clarke, L. L. (2005). Distributed security: Preventing cybercrime, 23 J. Marshall J. Computer & Info. L. 659. <https://repository.jmls.edu/jitpl/vol23/iss4/1/>
- Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., . . . Nic Lughadha, E. M (2015). Green plants in the red: A baseline global assessment for the IUCN sampled Red List Index for Plants. *PLOS ONE*, 10, Article e0135152.
- Cadwalladr, C., & Graham-Harrison, E. (2018, March 18). The Cambridge analytica files. *The Guardian*. http://davelevy.info/Downloads/cabridgeanalyticafiles%20-the-guardian_20180318.pdf
- Carlson, R. V., Boyd, K. M., & Webb, D. J. (2004). The revision of the Declaration of Helsinki: Past, present and future. *British Journal of Clinical Pharmacology*, 57(6), 695–713.
- Chan, H. K., Zhang, H., Yang, F., & Fisher, G. (2015). Improve customs' systems to monitor global wildlife trade. *Science*, 348(6232), 291–292.
- Chan, J., & Bennet Moses, L. (2016). Is big data challenging criminology? *Theoretical Criminology*, 20(1), 21–39.
- Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York University Press.
- CITES. (2016). *Wildlife crime. What are the challenges to effective law enforcement?* <https://cites.org/prog/icwcw.php/Wildlife-Crime>
- CITES. (2017). *CITES-listed medicinal plant species* (PC23 Inf. 10). <https://cites.org/sites/default/files/eng/com/pc/23/inf/E-PC23-Inf-10.pdf>
- Coudert, F. (2010). When video cameras watch and screen: Privacy implications of pattern recognition technologies. *Computer Law & Security Review*, 26(4), 377–384.
- Cowan, R. S., Chase, M. W., Kress, W. J., & Savolainen, V. (2006). 300,000 species to identify: Problems, progress, and prospects in DNA barcoding of land plants. *TAXON*, 55(3), 611–616.
- Décary-Héту, D., & Aldridge, J. (2015). Sifting through the net: Monitoring of online offenders by researchers. *The European Review of Organised Crime*, 2(2), 122–141.
- Demant, J., Munksgaard, R. D., écary-Héту, D., & Aldridge, J. (2018). Going local on a global platform: A critical analysis of the transformative potential of cryptomarkets for organized illicit drug crime. *International Criminal Justice Review*, 28(3), 255–274.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*.
- Di Minin, E., Fink, C., Tenkanen, H., & Hippala, T. (2018). Machine learning for tracking illegal wildlife trade on social media. *Nature Ecology & Evolution*, 2, 406–407.
- Dobson, A. P. (1996). *Conservation and biodiversity*. Princeton University Press.
- Elliott, L. (2012). Fighting transnational environmental crime. *Journal of International Affairs*, 66(1), 87–101.
- Engler, M., & Parry-Jones, R. (2007). *Opportunity or threat: The role of the European Union in global wildlife trade*. TRAFFIC Europe.
- Ess, C. (2002). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee*. <https://aoir.org/reports/ethics.pdf>
- EU Commission. (2016a). *EU action plan against wildlife trafficking*. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2016:87:FIN>
- EU Commission. (2016b). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

- Fajardo del Castillo, T. (2016). The European Union's approach in the fight against wildlife trafficking: Challenges ahead. *Journal of International Wildlife Law and Policy*, 19(1), 1–21.
- Felbab-Brown, V. (2017). *The extinction market: Wildlife trafficking and how to counter it*. C. Hurst & Co Publishers.
- Ferguson, A. G. (2012). Predictive policing and reasonable suspicion. *Emory Law Journal*, 62, 259–325.
- Flick, C. (2016). Informed consent and the Facebook emotional manipulation study. *Research Ethics*, 12(1), 14–28.
- Franklin, J., Paxson, V., Savage, S., & Perrig, A. (2007, October 29–November 2). *An inquiry into the nature and causes of the wealth of internet miscreants* [Conference session]. *ACM Conference on Computer and Communications Security (CCS)*, Alexandria, VA, United States.
- Gibbs, C., Gore, M. L., McGarrell, E. F., & Rivers, L. (2010). Introducing conservation criminology. Towards interdisciplinary scholarship on environmental crimes and risks. *The British Journal of Criminology*, 50(1), 124–144.
- Goetsch, B., Hilton-Taylor, C., Cruz-Piñón, G., Duffy, J. P., Frances, A., Hernández, H. M., Inger, R., . . . Gaston, K. J. (2015). High proportion of cactus species threatened with extinction. *Nature Plants*, 1(10), Article 15142.
- Grabosky, P. N., & Smith, R. G. (2001). Telecommunication fraud in the digital age: The convergence of technologies. In D. S. Wall (Ed.), *Crime and the Internet* (pp. 29–43). Routledge.
- Haas, T. C., & Ferreira, S. M. (2015). Federated databases and actionable intelligence: Using social network analysis to disrupt transnational wildlife trafficking criminal networks. *Security Informatics*, 4, Article 2.
- Haenlein, C., & Smith, M. L. R. (2017). *Poaching, wildlife trafficking and security in Africa: Myths and realities*. RUSI Publications.
- Harrison, J. R., Roberts, D. L., & Hernandez-Castro, J. (2016). Assessing the extent and nature of wildlife trade on the dark web. *Conservation Biology*, 30(4), 900–904.
- Head, L., Atchison, J., Phillips, C., & Buckingham, K. (2014). Vegetal politics: Belonging, practices and places. *Social and Cultural Geography*, 15(8), 861–870.
- Herbig, F. J. W., & Joubert, S. J. (2006). Criminological semantics. Conservation criminology—Vision or vagary? *Acta Criminologica*, 19(3), 88–103.
- Hinsley, A., de Boer, H. J., Fay, M. F., Gale, S. W., Gardiner, L. M., Gunasekara, R. S., Kumar, P., . . . Phelps, J. (2017). A review of the trade in orchids and its implications for conservation. *Botanical Journal of the Linnean Society*, 186(4), 435–455.
- Hinsley, A., Lee, T. E., Harrison, J. R., & Roberts, D. L. (2016). Estimating the extent and structure of trade in horticultural orchids via social media. *Conservation Biology*, 30(5), 1038–1047.
- Hinsley, A., Nuno, A., Ridout, M., St John, F. A. V., & Roberts, D. L. (2017). Estimating the extent of CITES non-compliance among traders and end-consumers; Lessons from the global orchid trade. *Conservation Letters*, 10(5), 602–609.
- Hinsley, A., & Roberts, D. L. (2018). The wild origin dilemma. *Biological Conservation*, 217, 203–206.
- Holt, J. T. (2007). Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior*, 28(2), 171–198.

- Holt, T. J. (2017). Identifying gaps in the research literature on illicit markets online. *Global Crime*, 18(1), 1–10.
- Holt, T. J., & Lampke, E. (2010). Exploring stolen data markets online: Products and market forces. *Criminal Justice Studies*, 23(1), 33–50.
- Holt, T. J., Smirnova, O., & Chua, Y. T. (2016). *Data thieves in action: Examining the international market for stolen personal information*. Palgrave.
- International Fund for Animal Welfare. (2008). *Criminal nature: The global security implications of the illegal wildlife trade*.
- IFAW. (2017). *Disrupt: Wildlife cybercrime*.
- International Union for Conservation of Nature. (2018). *IUCN Red List of threatened species* [Summary statistics]. <https://www.iucnredlist.org/resources/summary-statistics>
- Interpol. (2013). *Project web: An investigation into the ivory trade over the Internet within the European Union*.
- Interpol. (2018). *Global wildlife enforcement: Strengthening law enforcement cooperation against wildlife crime*.
- Kretser, H. E., Wong, R., Robertson, S., Pershyn, C., Huang, J. M., Sun, F., Kang, A., & Zahler, P. (2015). Mobile decision-tree tool technology as a means to detect wildlife crimes and build enforcement networks. *Biological Conservation*, 189, 33–38.
- Krigas, N., Menteli, V., Chrysanthou, P., & Vokou, D. (2017). The electronic trade in endemic plants of Cyprus through the Internet. *Plant Biosystems*, 151(3), 387–393.
- Lavorgna, A. (2014). Wildlife trafficking in the internet age. *Crime Science*, 3(5), 1–12.
- Lavorgna, A. (2020). *Cybercrimes: Critical issues in a global context*. Palgrave.
- Lavorgna, A., & Di Ronco, A. (2017). Fraud victims or unwary accomplices? An exploratory study of online communities supporting quack medicine. In P. C. van Duyne, J. Harvey, G. A. Antonopoulos, & K. von Lampe (Eds.), *The many faces of crime for profit and ways of tackling it* (pp. 297–324). Wolf Legal Publishers.
- Lavorgna, A., Rutherford, C., Vaglica, V., Smith, M. J., & Sajeva, M. (2018). CITES, wild plants, and opportunities for crime. *European Journal on Criminal Policy and Research*, 24, 269–288.
- Lemieux, A. M. (2014). *Situational prevention of poaching*. Routledge
- Lemieux, A. M. (2015). Geotagged photos: A useful tool for criminological research? *Crime Science*, 4, Article 3.
- Link. (2016). *Wildlife crime in 2016: A report on the scale of wildlife crime in England and Wales*. Wildlife and Countryside Link and Wales Environment Link.
- Mann, D., & Sutton, M. (1998). Netcrime. More change in the organization of thieving. *British Journal of Criminology*, 38(2), 201–229.
- Margulies, J. D., Bullough, L. A., Hinsley, A., Ingram, D., Cowell, C., Goettsch, B., Klitgård, B., . . . Phelps, J. (2019). Illegal wildlife trade and the persistence of “plant blindness.” *Plants People Planet*, 1, 173–182.
- Margulies, J. D., Hinsley, A., & Phelps, J. (2018). Illegal wildlife trade endangers plants—But few are listening. *IUCN Blog*. <https://www.iucn.org/news/species/201810/illegal-wildlife-trade-endangers-plants-few-are-listening>
- Markham, A., & Buchanan, E. (2012). *Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee (version 2.0)*. <https://aoir.org/reports/ethics2.pdf>
- McGough, H. N., Groves, M., Mustard, M., & Brodie, M. (2004). *CITES and plants: A user's guide version 3.0*. Royal Botanic Gardens, Kew.

- Michael, K., McNamee, A., & Michael, M. G. (2006, July 25–27). The emerging ethics of human-centric GPS tracking and monitoring [Conference session]. *2006 International Conference on Mobile Business, Copenhagen, Denmark*.
- Miller, C., Pay, J., & Smith, J. (2019). *Detecting online environmental crime markets* [Digital Dangers Report]. <https://globalinitiative.net/detecting-online-environmental-crime-markets/>
- Moreto, W. D., & Lemieux, A. M. (2015). From CRAVED to CAPTURED: Introducing a product-based framework to examine illegal wildlife markets. *European Journal on Criminal Policy and Research, 21*, 303–320.
- Moshier, A., Steadman, J., & Roberts, D. L. (2019). Network analysis of a stakeholder community combatting illegal wildlife trade. *Conservation Biology*. Advance online publication. <https://doi.org/10.1111/cobi.13336>
- Novoa, A., Le Roux, J. J., Richardson, D. M., & Wilson, J. R. (2017). Level of environmental threat posed by horticultural trade in Cactaceae. *Conservation Biology, 31*(5), 1066–1075.
- Olmos-Lau, V. R., & Mandujano, M. C. (2016). An open door for illegal trade: Online sale of strombocactus disciformis (cactaceae). *Nature Conservation, 15*, 1–9.
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data increases inequality and threatens democracy*. Crown.
- Phelps, J., & Webb, E. L. (2015). “Invisible” wildlife trades: Southeast Asia’s undocumented illegal trade in wild ornamental plants. *Biological Conservation, 186*, 296–305.
- Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Presa Reyes, M., Shyu, M. L., Chen, S. C., & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Survey, 51*(5), Article 92.
- Royal Botanic Gardens, Kew. (2017). *State of the world’s plants*.
- Regan, K. E. (2004). The need for a comprehensive approach to protecting rare plants: Florida as a case study. *Natural Resources Journal, 44*, 12 5–162.
- Roberts, D. L., & Hernandez-Castro, J. (2017). Bycatch and illegal wildlife trade on the dark web. *Oryx, 51*(3), 393–394.
- Runhovde, S. R. (2017). Taking the path of least resistance? Decision-making in police investigations of illegal wildlife trade. *Policing: A Journal of Policy and Practice, 11*(1), 87–102.
- Sajeva, M., Augugliaro, C., Smith, M. J., & Oddo, E. (2013). Regulating internet trade in CITES species. *Conservation Biology, 27*(2), 429–430.
- Sajeva, M., & Carimi, F. (1994). *Il ruolo della micropropagazione nella conservazione di piante succulente minacciate di estinzione per la loro importanza economica* [The role of micropropagation for the conservation of endangered succulents and their economic importance]. *Quaderni di Botanica Ambientale e Applicata, 5*, 55–60.
- Sajeva, M., Carimi, F., & McGough, N. (2007). The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and its role in conservation. *Functional Ecosystems and Communities, 1*(2), 80–85.
- Schippmann, U., Leaman, D. J., & Cunningham, A. B. (2002). *Biodiversity and the ecosystem approach in agriculture, forestry and fisheries. Impact of cultivation and gathering of medicinal plants on biodiversity: Global trends and issues*. FAO.
- Shapiro, A. (2017). Reform predictive policing. *Nature, 541*(7638), 458–460.
- Smirnova, A., & Cudré-Mauroux, P. (2018). Relation extraction using distant supervision: A survey. *ACM Computing Surveys, 51*(5), Article 106.
- Sollund, R. (2015). Wildlife trafficking in a globalised world: An example of motivations and modus operandi from a Norwegian case study. In F. Angelici (Ed.), *Problematic wildlife* (pp. 553–570). Springer.

- Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-Jensen, A., Scholtens, I., Peelen, T., Prins, T. W., & Kok, E. (2016). Advances in DNA metabarcoding for food and wildlife forensic species identification. *Analytical and Bioanalytical Chemistry*, 408, 4615–4630.
- Steinmetz, K. F. (2016). *Hacked: A radical approach to hacker culture and crime*. New York University Press.
- Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical challenges in online research: Public/private perceptions. *Research Ethics*, 13(3–4), 184–199.
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, 7(3), 321–326.
- The United Nations Office on Drugs and Crime. (2016). World wildlife crime report: Trafficking in protected species.
- Van Den Hoonaard, W. C. (2013). The social and policy contexts of the New Brunswick Declaration on Research Ethics, Integrity, and Governance: A commentary. *Journal of Empirical Research on Human Research Ethics*, 8(2), 104–109.
- van Uhm, D. P., & Wong, R. W. Y. (2018). Establishing trust in the illegal wildlife trade in China. *Asian Journal of Criminology*, 14(1), 23–40.
- Wandersee, J. H., & Schussler, E. E. (1999). Preventing plant blindness. *The American Biology Teacher*, 61(2), 82–86.
- Ward, K. (2018). Social networks, the 2016 US presidential election, and Kantian ethics: Applying the categorical imperative to Cambridge Analytica's behavioral microtargeting. *Journal of Media Ethics*, 33(3), 133–148.
- Williams, M., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *British Journal of Criminology*, 57(2), 320–340.
- Wu, J. (2007). World without borders: Wildlife trade on the Chinese-language internet. *TRAFFIC Bulletin*, 21(2), 75–84.
- World Wildlife Fund. (2016). *CITES CoP17: Summary of WWF's positions on priority agenda items*.
- World Wildlife Fund. (2018). *Living planet report 2018: Aiming higher* (M. Grooten & R. E. A. Almond Eds.).
- Wyatt, T. (2013). *Wildlife trafficking: A deconstruction of the crime, the victims and the offenders*. Palgrave.
- Young, S. (2003). Contemporary issues of the convention on international trade in endangered species of wild Fauna and Flora (CITES) and the debate over sustainable use. *Journal of International Environmental Law and Policy*, 14(1), 167–189.
- Zimmerman, M. E. (2003). The black market for wildlife: Combating transnational organized crime in the illegal wildlife trade. *Vanderbilt Journal of Transnational Law*, 36, 1657–1689.

Author Biographies

Anita Lavorgna is an associate professor in criminology within the Department of Sociology, Social Policy & Criminology at the University of Southampton. She is currently working on research projects on internet-facilitated wildlife trafficking, health frauds, and cybercrimes.

Stuart E. Middleton is a lecturer in computer science within the Department of Computer Science at the University of Southampton. He is currently working on research projects involving Natural Language Processing and Information Extraction of wildlife trafficking and cyber-crime intelligence from online forums/marketplaces, climate change observations from historical records, and legal risk markers from the UK's HM Land Registry title deed documents.

Brian Pickering is a senior research fellow in electronics and computer science at the University of Southampton. He is currently working on ethical research into online and big data, and the tension between privacy laws and ethical expectations.

Geoff Neumann is a research fellow in computer science at the University of Southampton. He is currently working on natural language processing and web crawling in projects related to wildlife crime and risk analysis in land registry documents.