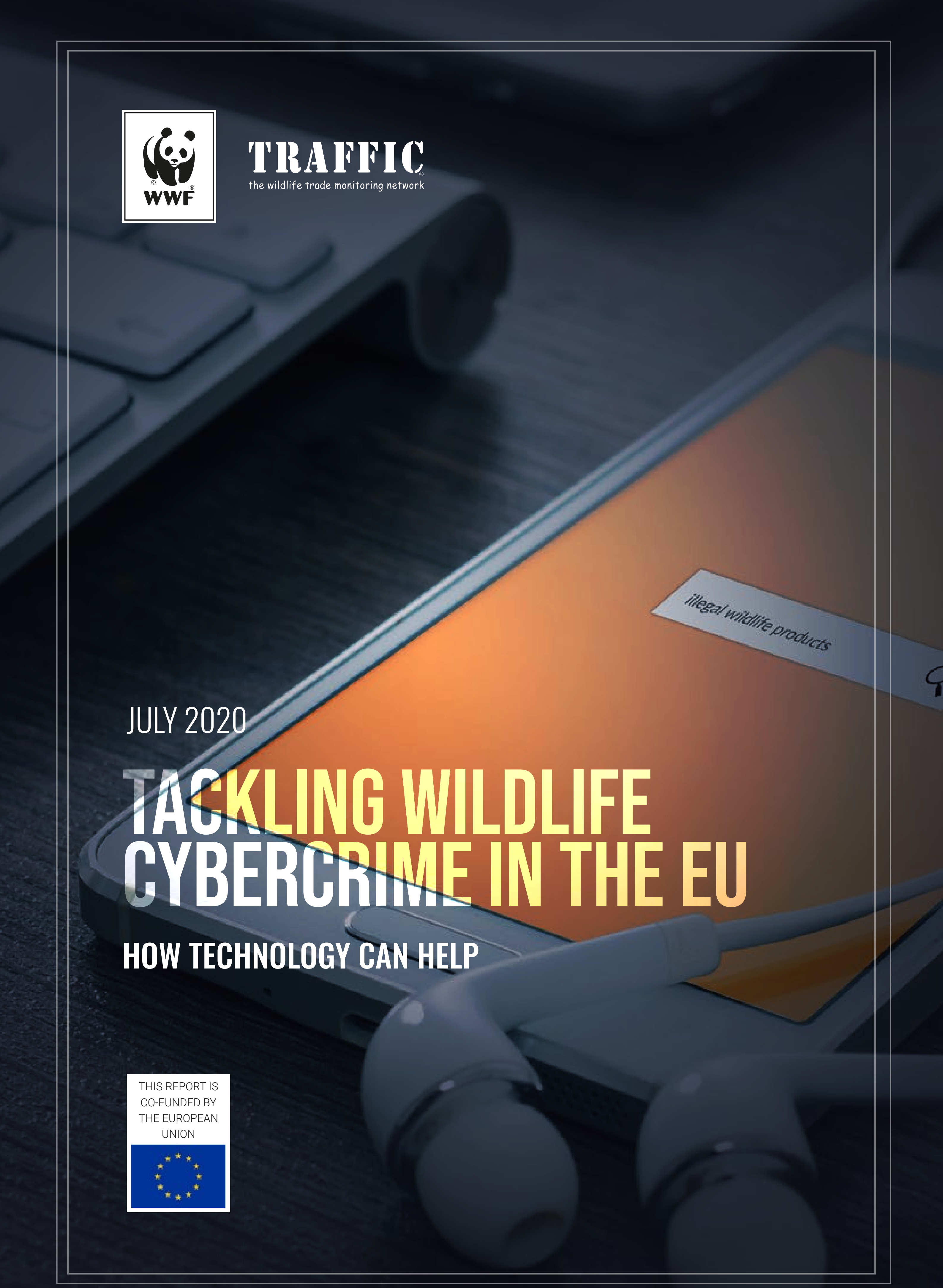


WWF

TRAFFIC
the wildlife trade monitoring network

JULY 2020

TACKLING WILDLIFE CYBERCRIME IN THE EU

HOW TECHNOLOGY CAN HELP

THIS REPORT IS CO-FUNDED BY THE EUROPEAN UNION

JOINT REPORT

This report is published by WWF-Belgium with in-kind support from TRAFFIC in the context of the EU Wildlife Cybercrime project, an EU-funded initiative aiming to disrupt and deter criminals and their networks trafficking wildlife in, or via, the EU using the internet and parcel delivery services. This project is implemented by WWF, IFAW, INTERPOL and the Belgian Customs, with support from TRAFFIC.

Project webpage:
https://wwf.be/fr/wildlife-cybercrime/

This report is targeted primarily at law enforcement authorities in the European Union involved in tackling wildlife cybercrime.

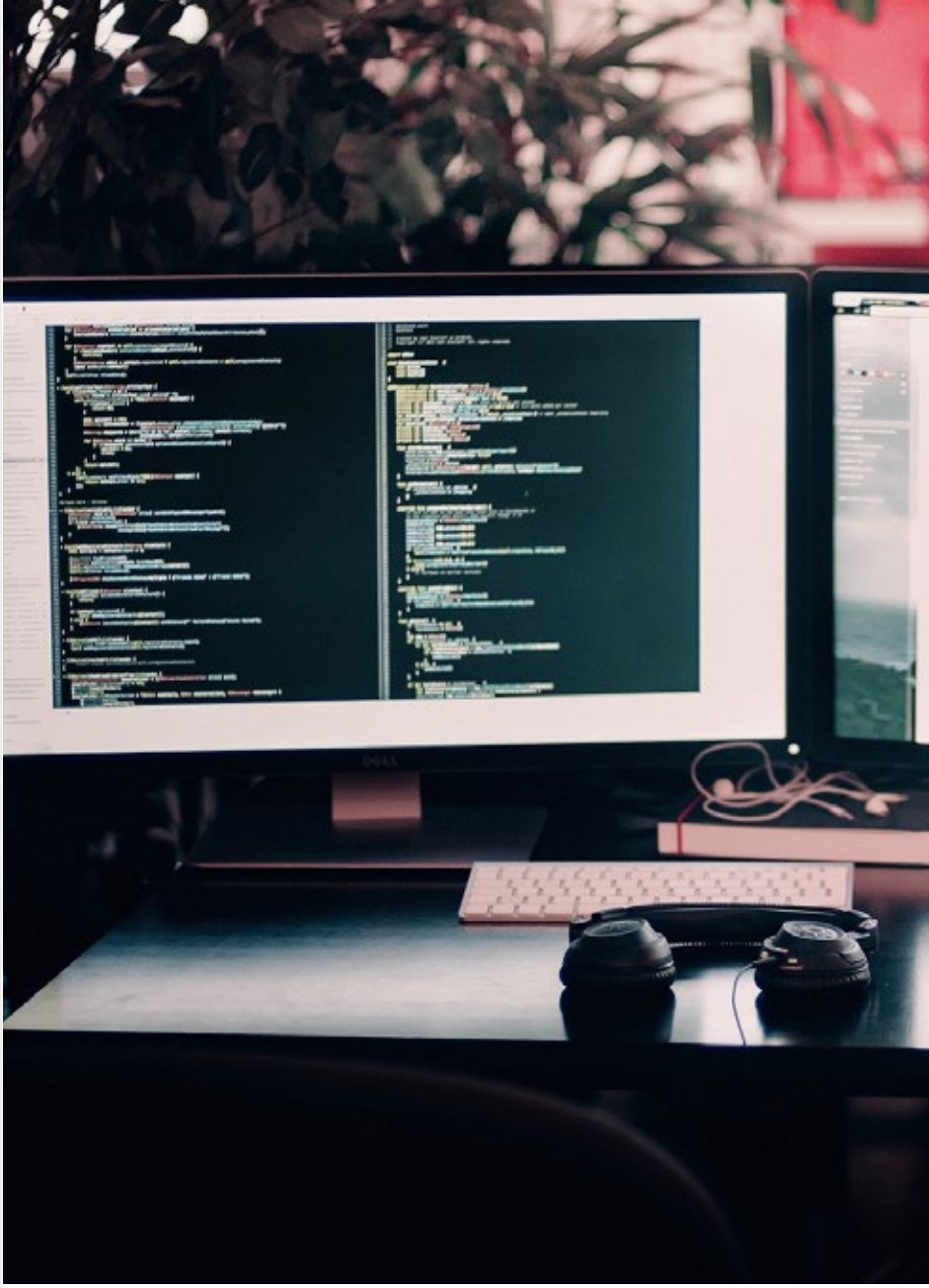
WWF and TRAFFIC do not endorse the use of any of the tools examined in this report nor can they vouch for their accuracy/functionality. The list of tools reviewed is not exhaustive and no compensation, financial or otherwise, was offered or received for any of the products reviewed.

This report was funded by the European Union's Internal Security Fund — Police. The content of this report represents the views of the authors only and is their sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

LEAD AUTHOR
Florian Debève, TRAFFIC/WWF
LEAD SUPERVISOR
Emilie Van der Henst, TRAFFIC/WWF
CONTRIBUTORS AND PEER REVIEWERS
David Roberts (University of Kent), Enrico Di Minin (University of Helsinki), Vincent Danjean (INTERPOL), Katalin Kecse-Nagy (TRAFFIC), Lu Gao (TRAFFIC), Crawford Allan (TRAFFIC) & Floris van de Gevel (Netherlands Food and Consumer Product Safety Authority).

PUBLISHED BY:
WWF-Belgium,
Bd E. Jacqmain 90, 1000 Brussels.
E.R.: Antoine Lebrun

DESIGN
Marcus Cornthwaite



ACKNOWLEDGEMENTS

The author thanks David Roberts from the University of Kent, Enrico Di Minin from the University of Helsinki, Jenifer Jacquet from the University of New York, Joss Wright from the University of Oxford, Vincent Danjean from INTERPOL, Floris van de Gevel from the Netherlands Food and Consumer Product Safety Authority (NVWA), Wim Vandeveld from Leuven University, and Hasita Bhammar from the World Bank, for their contribution and support to the study. Thanks also go to WWF and TRAFFIC colleagues Emilie Van der Henst, Katalin Kecse-Nagy, Lu Gao, Crawford Allan, Vinciane Sacré, Antony Bagott, Marcus Cornthwaite, Yu Xiao, Ling Xu, Weihua Xin and IFAW colleague Mia Crnojevic, for their ongoing support with project design and report reviews. Finally, the European Union is gratefully thanked for its financial support of this study.

TABLE OF CONTENTS

EXECUTIVE OVERVIEW

EXECUTIVE SUMMARY

The growing use of online marketplaces for wildlife trafficking has been established by several studies over the last decade. Such trafficking presents unprecedented challenges to law enforcement investigations, particularly given the number, speed and anonymity of transactions that Internet-based platforms enable.

Law enforcement agencies (LEAs) and non-governmental organisations (NGOs) still rely largely on manually collecting data from the internet for identifying wildlife cybercrime[1], a very labour and time-intensive exercise. Technology is often regarded as a potential game-changer in this area. There is an untapped potential for automated data analytical tools to provide relevant information, with promising developments from a range of academic institutions, private companies, civil society and law enforcement authorities. Beyond accelerating and improving their monitoring activities, these tools could help practitioners to draw up a more comprehensive picture of the true scale and nature of wildlife cybercrime in the EU, including identification of where it occurs on the internet and which species are most affected. Developing this picture is important for supporting investigations, for governments to develop a sense of how best to respond, and for effective engagement with the online technology companies whose platforms are unwittingly being used.

This report offers a general assessment of how technology can support LEAs in their daily efforts to address wildlife cybercrime. Section 1 includes a series of key definitions, as necessary background to understand the potential of the different types of solutions discussed. Sections 2 and 3 provide a non-exhaustive overview of a series of advanced tools that could help law enforcement officials to counter wildlife cybercrime.

In order to gather the information presented in this report, a qualitative method was followed, largely based on a dozen interviews with experts and leading authors on the subject, as well as on a broad review of the existing literature. The goal of this report is not to provide an exhaustive presentation of all the solutions that currently exist, but rather to showcase those that our consultations and literature review enabled us to identify, in order to understand how technology can help address wildlife cybercrime.

SOLUTIONS

Section 2 presents a series of tools that can provide solutions to enable LEAs to exploit the potential of data analytical techniques to address wildlife cybercrime through accessing and analysing a large amount of online data, with the potential greatly to reduce the human effort needed to monitor illegal wildlife trade online. They encompass several techniques related to data and computer sciences, such as automated web scraping, data mining, machine learning, natural language processing and image recognition.

The study finds that despite considerable progress, none of the solutions presented in section 2 provides the silver bullet for addressing wildlife cybercrime.

[1] The term "wildlife cybercrime" is used throughout the report as a synonyms for other accepted terminologies such as "wildlife trafficking online" and "wildlife crime linked to the internet".

The solutions and associated tools are presented under three categories:

1 NON-COMMERCIAL SOLUTIONS DEVELOPED BY ACADEMIA AND CIVIL SOCIETY

All the tools presented here have been developed with the clear intention of being applied to wildlife trafficking in order to support LEAs. Some of them are the result of close collaboration between LEAs and universities. They show convincing results when applied to a small number of species or products, and on a limited number of platforms. However, despite these encouraging results, none of these tools has yet been tested on a larger scale, their biggest common drawback.

2 COMMERCIAL SOLUTIONS WHICH CAN POTENTIALLY BE TAILORED TO THE NEEDS OF LEAS

The number of private companies providing commercial services for web-scraping, data mining or machine learning is growing. Such services are largely tailored to commercial purposes and used by companies as a competitive intelligence tool. However, such tools could potentially be adapted and used to monitor wildlife trafficking online. Engaging an external company for this purpose could be an option for LEAs. Our research shows that the use of commercial tools for monitoring wildlife trafficking online is relatively limited to date.

Those we interviewed said there were three main reasons for this. First, there are often complexities around procurement regulations, including company certification rules when public bodies work alongside companies. Second, company collaborations are usually more expensive than those with the academic world. Third, flexibility, responsiveness and working conditions are described as more favourable when collaborating with academic researchers.

3 IN-HOUSE SOLUTIONS AND OTHER OPTIONS

Developing software in-house can guarantee a more tailored approach to the specific needs of LEAs and avoids externalising sensitive information or key processing methodology with external third parties. In some cases, it can reduce procurement and administrative costs. A number of alternative options were also examined. They fall into the following categories:

Communication and collaboration: communication tools between LEAs could play a two-fold role: first in managing, sharing and processing large volumes of data often held by different agencies in order to create critical datasets; and, second, in improving the ability to report wildlife cybercrime. This section lists a series of tools, mostly institutional, developed to facilitate communication between agencies involved in the detection and investigation of wildlife crime.

Identification: this section highlights several initiatives to support the identification of protected species, including through phone apps. Some offer the ability to scan directly with the camera of a mobile phone, while others can deduce species based on a series of questions.

Investigation: several processing tools inspired by research into big data analysis techniques are aiming to simplify the task of investigators. Although online investigation tools are largely familiar to law enforcement agencies, it is nevertheless useful to highlight a few in the context of wildlife cybercrime investigations.

Socialisation process: conducting illicit trade requires a customer base that is socialised into the norms, practices and practicalities of the trade. Some claim that the Internet is a powerful vehicle for creating new demand by socialising people into illegal trade. Research has shown that online platforms allow global demand opportunities to be opened up, turning what might have previously been a small interest group into a large global community of potential buyers. This process is reinforced by a certain sense of impunity, inherent to wildlife cybercrime due to a general lack of enforcement. Some private sector companies have implemented measures to deter illegal trade, such as the pop-up notifications that appear on Instagram when users search for animals threatened by illegal trade, or the recent announcement by Tencent that it will introduce a reporting function for users to flag suspected illegal activity.

Other: this section presents a variety of tools deemed useful but that do not fall under the above categories. These range from a search engine specifically designed for the dark web to crowdsourcing data processing, or "one-click" data collecting tools.

CONCLUSIONS

In recent years, computer science has moved forward rapidly and mastered algorithmic models that are able, via data mining and machine learning, of making understandable sets of data that would otherwise be too vast or too varied to be comprehended. Thanks to these tools, today it is possible not only to identify wildlife sold on the internet but also to assess illegality. Moreover, tools derived from data science have the ability to shed light on existing links and tendencies online, and also to anticipate and predict future trends. These types of tools could yield greater information about the dynamics of illicit markets, the nature of the networks trading online, and ways to understand consumer demand..

However, this report shows that it would be an illusion to think that technology alone, despite its appealing possible benefits, would be the ultimate means to counter wildlife cybercrime. At present, there is no one-fit-for-all, scalable, reliable, systematic and repeatable way to detect wildlife cybercrime automatically. Even though some tools reach a high level of accuracy when distinguishing between legal and illegal goods, their application is still restricted to a limited number of platforms (e.g. eBay) and products (e.g. ivory). Assessing automatically the illegality of an online advertisement selling wildlife still presents a major challenge due to primarily: the complexity of the wildlife trade regulations, covering many species and with several trade exemptions; and the lack of data available to "train" computers on a larger scale. Human expertise therefore remains key in this field.

Automatic systems still generate a lot of irrelevant results, which demand substantial, often overwhelming, cleaning efforts from law enforcement officers. While technology appears potentially capable of simplifying and accelerating monitoring tasks, it appears that today it is rather changing the nature of these tasks. The large amount of time usually needed for manually monitoring the web is now spent cleaning up the results collected through automated monitoring and scraping.

Furthermore, a critical challenge relates to the fact that both data mining and machine learning rely on the analysis of critical datasets. The amount of data needed to enable computers to locate and identify wildlife cybercrime is enormous and difficult to compile. Concerned entities amongst the LEAs and NGOs are unlikely to have the volume of data required, nor the technical means of collecting them. Specialised companies that have the means to acquire and process large amounts of data would appear to be best option.

Ultimately, using the highest potential of automated solutions to combat wildlife cybercrime will require a large multi-stakeholder collaboration and co-ordination involving all relevant stakeholders, such as police, customs services (environment and cyber units), academia, NGOs, and the giant tech-companies.

RECOMMENDATIONS

Based on the research conducted in the context of this report, the LEAs are recommended to:

Build and centralise the critical datasets necessary for scaling-up automated solutions

Strengthen the collaboration between law enforcement agencies on sharing best practices and knowledge on wildlife cybercrime

Strengthen the internal collaboration between the cyber units and wildlife crime units at the national level in order to tackle efficiently the multidisciplinary aspect of wildlife cybercrime

Strengthen internal skills and build capacity amongst the LEAs to facilitate the integration of technological solutions into the daily activities of officers responsible for monitoring wildlife crime

Strengthen the collaboration with major players in the private sector active in data collection and analysis to develop efficient tools and conduct decisive actions

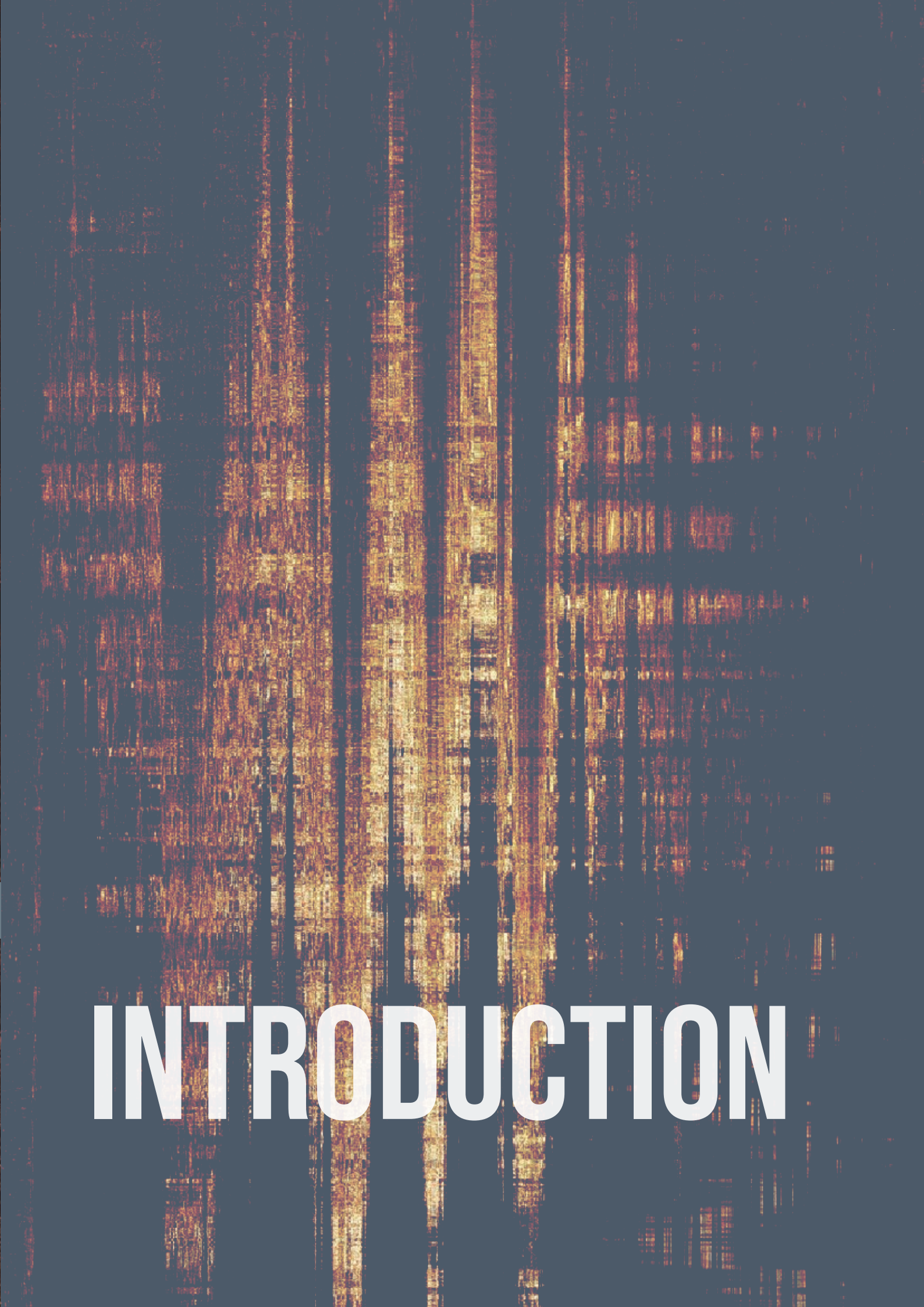
Strengthen the collaboration with representatives of civil society (NGOs and academia) to develop tailor-made tools

OVERVIEW OF THE TOOLS

NAME	DEVELOPER	IN BRIEF	TAXA OF FOCUS	AVAILABILITY / COST	STATE OF DEVELOPMENT	MORE INFORMATION
NON-COMMERCIAL TOOLS						
iTrade	University of Kent	This tool uses machine learning to differentiate between legal and illegal goods with a high level of accuracy.	Tested for ivory	Free of charge (available via publication) for LEAs	Product still under development. Developers seeking collaboration to improve the product	https://peerj.com/articles/cs-10/
San Diego models	University California	This model combines machine learning and Natural Language Processing (NLP) to successfully detect illegal content on Twitter without the use of pre-existing—manually built—training datasets. This could be key in overcoming the challenge of building data sets.	All taxa	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	https://www.frontiersin.org/articles/10.3389/fdata.2019.00028/full
WildTrade	University of Helsinki	This project proposes a comprehensive approach, combining computer vision and natural language processing methods to analyse concomitantly the text and in the visual content of digital platforms.	All taxa	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	NA
WildEye	New York University	It collects data from a large number of online retail and social media sites to identify illegal species being offered for sale, based on a series of keywords.	Tested on Annex-I listed species, but could be extended	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	NA
The Dynamic Data Discovery Engine	The Global Initiative Against Transnational Organized Crime	This is a methodology that aims to build as comprehensive a picture as possible of how, where and when plants and animals listed in CITES, or commodities containing them, are transacted over the internet. It consists of a large-scale analysis of how wildlife trade is discussed on the internet. This process also sets out to find keywords and key phrases that best identify wildlife cybercrime activity.	All taxa	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	https://globalinitiative.net/wp-content/uploads/2019/01/TGIATOC-DetectingOnlineMarkets-Web-1.pdf
ChimpFace	Conservation X Labs	This image recognition software has been able to successfully recognise species as well as individual Chimpanzees based on photos published on the internet.	Great Apes	NA	Product still under development. Developers seeking collaboration to improve the product	https://conservationx.com/project/id/8

NAME	DEVELOPER	IN BRIEF	TAXA OF FOCUS	AVAILABILITY / COST	STATE OF DEVELOPMENT	MORE INFORMATION
TITANIUM project	INTERPOL	TITANIUM researches, develops and validates novel data-driven techniques and solutions designed to support LEAs investigating criminal or terrorist activities involving virtual currencies and/or underground markets on the darknet.	All taxa	Free of charge for INTERPOL member states	Finished products.	https://dws.pm/monitor/ https://graphsense.info/
COMMUNICATION AND COLLABORATION TOOLS						
EU-TWIX	Initially developed by Belgian police, managed by TRAFFIC	It enables near real-time exchange of information between enforcement and management officials. Types of information shared include seizure details such as modus operandi, stolen specimen alerts and assistance requests for species identification.	All taxa	Free of charge for LEAs and CITES management authorities	Finished product. Developers seeking contribution.	https://www.eu-twix.org/
CENcomm (ENVIRONET)	World Customs Organisation	It allows Customs officers and other authorized users to exchange real-time information during an operation. It also provides a means to share information in a standardized format.	All taxa	Free of charge for members of WCO	Finished product. Developers seeking contribution.	http://www.wcoomd.org/en/topics/enforcement-and-compliance/instruments-and-tools/cen-suite/cencomm.aspx
IDENTIFICATION TOOLS						
Wildscan Mobile App	USAID	The application makes it possible for everyone, including citizens, to easily identify and report wildlife trafficking and support ongoing law enforcement efforts to counter wildlife crime.	All taxa	Free access	Finished product.	https://wildscan.apk.cafe/
Wildlife Alert Mobile App	The Wildlife Conservation Society's	It helps users to assess the authenticity and the source of wildlife items. If an item is thought to be from a protected species, the user is provided with instructions on how to proceed.	All taxa	Free access	Finished product.	https://apps.wcswildlifetrade.org/
INVESTIGATION TOOLS						
Web-IQ	Private company	This tool offers reliable and customised open source intelligence (OSINT) investigation support	NA	Private company for profit	Finished product.	https://www.web-iq.eu/

NAME	DEVELOPER	IN BRIEF	TAXA OF FOCUS	AVAILABILITY / COST	STATE OF DEVELOPMENT	MORE INFORMATION
Anita Project	ENGINEERING Ingegneria Informatica S.p.A. (EU-funded project)	This project is working on delivering a set of advanced tools and techniques to efficiently address online illegal trafficking.	NA	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	https://www.anita-project.eu/
Nuix	Private company	Nuix provides an easy-to-use and collaborative digital forensics tool.	NA	Private company for profit	Finished product. Developers seeking collaboration.	https://www.nuix.com/
OTHER TOOLS						
The Global Database of Events, Language, and Tone (GDELT) project	Private developers	Monitors print, broadcast, and web news media in about 100 languages from across reportedly every country in the world.	NA	Free access	Finished product.	https://www.gdelt-project.org/
Aleph Search Engine	Private developers	Aleph is a search engine that references the dark and the deep web.	All taxa	Private company for profit	Finished product.	NA
One-Click Tool	University of Kent	The software can download the details of a suspicious webpage, in a few seconds, in an organised and consolidated fashion, thus saving a considerable amount of time that would otherwise be spent copying and pasting the text and downloading images.	All taxa	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product	Contact David Roberts, David Roberts, D.L.Roberts@kent.ac.uk
Ivory crowdsourcing tool	University of Kent	This tool uses crowdsourcing to identify elephant ivory and train computer to recognize images of ivory.	Ivory	Free of charge for LEAs	Product still under development. Developers seeking collaboration to improve the product.	Contact David Roberts, David Roberts, D.L.Roberts@kent.ac.uk



INTRODUCTION

INTRODUCTION

The growing use of online marketplaces for global wildlife trafficking has been established by several studies over the last decade (Lavorgna, 2014; Xiao & Wang, 2015; Yu, *et al*., 2017; IFAW, 2018; TRAFFIC, 2019). Yet, it remains a challenge to gauge the extent, nature and impact of online interactions in the illegal wildlife trade supply chain (Haysom, 2018).

Wildlife trafficking on Internet-based platforms, given the volume, pace, and anonymity that they allow, represent unprecedented obstacles to law enforcers' investigations. However, the urgency to address this issue has encouraged researchers and practitioners across the globe to consider the opportunities offered by the digital age and to develop tools that could support law enforcement agencies (LEAs), non-governmental organisations (NGOs) and other practitioners in tracking wildlife cybercrime and thus contribute to reducing online wildlife trafficking by making these activities harder and riskier for those involved.

LEAs and NGOs still rely largely on manually collecting data from the internet for identifying wildlife cybercrime-related posts; a very time-intensive exercise where the amount of data to monitor and analyse outpaces human capacity. There is an untapped potential for automated data analytic tools to provide this information, with promising developments in this direction from a range of academic institutions, private companies, civil society and law enforcement authorities.

Beyond creating a robust, repeatable way of detecting wildlife trafficking online, such tools could contribute to the drawing up of a comprehensive picture of the true scale of wildlife cybercrime, including identification of where it occurs on the internet and which species are most affected by it. Developing this picture is important for supporting investigations, for governments to develop a sense of how best to respond, and for effective engagement with the online technology companies whose platforms are implicated.

This report proposes a general assessment of where technology can support the LEAs in their daily fight against wildlife cybercrime. It brings together a series of tools that can be available to law enforcement officials in order to make these solutions and their potential known to the LEAs and other practitioners, and to trigger their interest in using or further developing them.

The purpose of this report is not to offer an exhaustive review of all these solutions but rather to focus on recent examples of advances in data and computer science, and tools that derive from them, to illustrate possibilities of automated monitoring of wildlife cybercrime.

1.1. KEY DEFINITIONS

Definitions are provided below of some of the core concepts useful to understand the tools and solutions presented in this report.

1.1.1. WEB SCRAPING

Scraping generally defines a technique for extracting content (information) from one or more websites—or social media—completely and automatically. Scripts are used to extract the searched information. In computer science, a script designates a program responsible for executing a pre-defined set of actions when a web page is displayed on a screen. It is a sequence of commands that allows the automation of successive tasks in a given order. In the case of scraping, the scripts organise the automated downloading of web content according to pre-defined queries (Butterfield, *et al*., 2016).

DEFINITION

Scraping can be used by commercial actors as a tool for monitoring competitors (e.g. automatically retrieving the rates charged by a competitor online and detecting their variations). Moreover, they are also technical tools used by police and criminal justice authorities in online research and investigations that could be applied to monitor and identify wildlife cybercrime.

Web scrapers can be used on online open sources—the open web (e.g. public Twitter or Instagram feeds) and the dark web—or restricted—the deep web (e.g. Facebook feeds), whose access is protected by passwords or other means of authentication. While the collection of data on the open web and the dark web does not a priori require any specific authorisation, as this information belongs to the public domain[2], the analysis of restricted sources usually requires prior judicial authorisation[3]. Although it also depends on the respective national legislation, it is imperative that the appropriate authorisations are sought, delivered and validated before starting a monitoring procedure based on automated web-scraping techniques.

Furthermore, since May 2018, the General Data Protection Regulation (GDPR) has applied in the Member States of the European Union. It is important to be aware of the potential interferences of scraping activities with this regulation. Under GDPR, to use or hold the personal data of any EU citizen, one must comply with one or more of the legal provisions foreseen for storing or exploiting personal data, without breaching the regulation. Article 9 of the GDPR[4] describes specific personal data categories with exemptions and therefore introduces the so-called "Police-Justice Directive"[5] which applies to law enforcement authorities.

Generally, under the GDPR, a business enterprise, an NGO or a LEA carrying out scraping activities must comply with one or more of the legal exemptions provided in the Regulation (EU) 2016/679. There are five of these, listed below.

- ✓ **THE CONSENT:** the owner of the data has consented to the use of their data. This argument could, for example, be used to justify the scraping of data accessible to the public on platforms such as Facebook or Instagram.
- ✓ **THE CONTRACT:** scraping personal data is necessary for the performance of a contract with the data subject. This will a priori not concern us in the context of this report.
- ✓ **THE COMPLIANCE:** scraping personal data is necessary for compliance with a legal obligation. This argument could be raised by law enforcement if this can be linked to their legal obligation to uphold the law.
- ✓ **THE VITAL INTEREST, PUBLIC INTEREST OR OFFICIAL AUTHORITY:** this is the argument generally applicable to bodies managed by the State when access to personal data is in the public interest.
- ✓ **THE LEGITIMATE INTEREST:** legitimate interests is more flexible and could in principle apply to any type of processing for any reasonable purpose. The GDPR does not have an exhaustive list of what purposes are likely to constitute a legitimate interest. However, it does say the following purpose constitute a legitimate interest: indicating possible criminal acts.

[2] Please note that being in the public domain means that one can scrape whatever/however she/he wishes to do so. Some ethical rules must, in some instances, be respected. Therefore, it depends on the ethics of the organisation conducting the data collecting operation.

[3] Please, refer to Article 32 of the Convention on Cybercrime, Ets 185. https://www.europarl.europa.eu/meetdocs/2014_2019/documents/libe/dv/7_conv_budapest_/7_conv_budapest_en.pdf

[4] The text of Regulation (EU) 2016/679 is available here https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[5] The "Police-Justice" Directive n° 2016/680, 27th April 2016 is available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32016L0680

Practitioners in search of more information on these legal issues can refer directly to Regulation (EU) 2016/679 or (EU)2016/680 and consult their relevant national bodies. James Wingard and Maria Pascual's paper Catch me if you can, published in 2018 by the Global Initiative Against Transnational Organized Crime (Wingard & Pascual, 2018) is also a useful reference for more information about the legal challenges to combat wildlife cybercrime.

1.1.2. DATA MINING

DEFINITION

Typically, data mining refers to the analysis of data from different perspectives and transforming those data into useful information, establishing relationships between datasets or identifying patterns (Butterfield, *et al***., 2016).**

Data mining software is part of the set of analytical tools used for data analysis. Mining software allows users to analyse data from different angles, categorise data and summarise identified relationships. Technically, data mining is the process of finding correlations or patterns between a certain number of related databases. Data mining relies on complex and sophisticated algorithms to segment data and evaluate future probabilities.

Data mining consists of making understandable a set of data that would otherwise be too vast or too varied to be comprehended. This not only has the ability to shed light on existing links and trends (in the case of wildlife cybercrime, that could help understanding some virtual routes, the most traded species, as well as some other useful information concerning, for example, the sociology of the people involved in the trafficking), but also to anticipate and predict future trends. This type of tool could yield greater information about the dynamics of illicit markets, the nature of the networks trading online, and automated ways to understand consumer demand or to differentiate between legal and illegal goods (Haysom, 2018).

1.1.3. ARTIFICIAL INTELLIGENCE

DEFINITION

Artificial Intelligence (AI) aims to allow machines, and more particularly computer systems, to simulate human cognitive processes.

These processes include machine learning (acquiring information and rules related to their use, see definition below), reasoning (applying the rules to reach approximate or precise conclusions) and self-correction. Specific applications of AI also include natural language analysis (see Natural Language Processing section below), voice recognition and/or even artificial vision

1.1.4. MACHINE LEARNING

DEFINITION

Machine learning is a modern science for discovering patterns and making predictions from data based on statistics, data mining, pattern recognition and predictive analysis (Butterfield, *et al***., 2016).**

Machine learning is very effective in situations where insights must be discovered from large and diverse datasets (Big Data). For the analysis of such data, it is much more efficient than traditional methods in terms of accuracy and speed. For example, based on information associated with a transaction such as amount and location, and historical and social data with the right systems in place, machine learning can detect potential fraud nearly instantly (Abbasi, 2012).

It is important not to confuse machine learning with the previously described data mining, which, to put it simply, consists of an extra step. Machine learning is a subset of AI, "trained" for the ability to learn about new datasets without being programmed with an

explicit source. In other words, it involves training the computer to react automatically in a certain number of given situations, so that it can then, alone, and without any further programming, react in unknown situations. In the context of wildlife cybercrime, an example would be to teach a computer to recognise ivory illegally sold, based on a set of known data[6], so that it can independently recognise illegal ivory in an unknown set of data. A key element is that both data mining and machine learning ultimately rely on the analysis of critical datasets. The definition of critical is vague and depends greatly on the type of situations for which one wants to train the computer, but the amount of data needed is enormous. It is, therefore, important to realise that harvesting, transmitting, sharing, and analysing these data are at the core of the fight against wildlife crime. Later in this report, some solutions that could help pull together critical datasets are provided.

1.1.5. NATURAL LANGUAGE PROCESSING (NLP)

DEFINITION

Natural Language Processing is the ability of a computer program to understand human language as it is spoken. Essentially, it can be used to interpret free text and make it analysable (Butterfield, *et al*., 2016). It is part of AI technologies.

NLP application development is difficult because traditionally computers are designed to receive instructions from humans in a precise, unambiguous and highly structured programming language, or with a limited number of voice commands clearly set. But the "human to human" discourse is not always precise, often ambiguous, and its linguistic structure may depend on many complex variables, including slang, regional dialects and the social context.

Despite the complexity of these analyses, NLP can now mimic human comprehension of words and sentences and allows Machine learning models to process large amounts of information before providing accurate answers to questions asked to them. In the context of wildlife cybercrime, these technologies can be used to review large amount of product descriptions, commentaries, and forum discussions. Some of the methodologies presented below are building on these models.

1.1.6. APPLICATION PROGRAMMING INTERFACE (API)

DEFINITION

Websites, or social media, are built using coding language. Most of the time these quite sophisticated codes represent an important investment, especially for large Internet platforms, which prefer to keep them proprietary in order to prevent others from freely exploiting their innovation. Essentially, this is contrary to open source communities, such as Linux, who operate under the philosophy that open access to codes, replicable and editable by all, is more likely to produce better products.

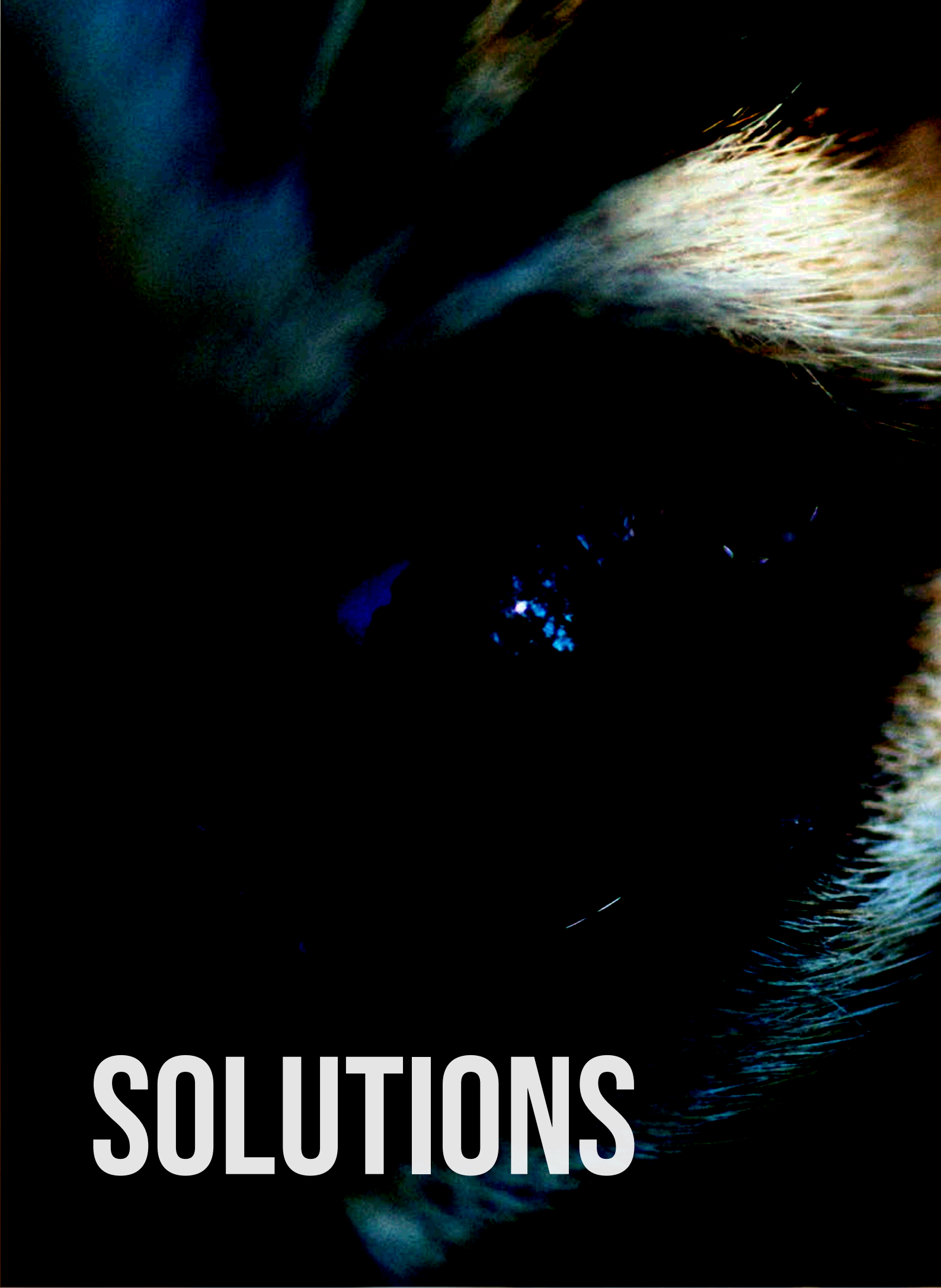
However, sometimes, a window opens on a limited part of a website's codes, because the developers are seeking to allow access to other applications, which could possibly enrich the overall website. These include, for example, games and applications created by independent developers that can be added by users to their Facebook feeds. In order to allow these applications to interact properly with its platform, Facebook had to give open access to a certain part of its programming codes, thus offering a way for other sites to connect. It is this way of allowing interaction between two applications that is called Application Programming Interface or API (Butterfield, *et al*., 2016). Some of the tools presented below have successfully leveraged the APIs of larger platforms (e.g. eBay's) to identify activities related to wildlife cybercrime.
.

[6] A dataset where legal and illegal ivory items have been previously sorted.

1.2. METHODS

In order to gather the information presented in this report, a qualitative method was followed largely based on a dozen interviews with experts and leading authors on the subject, as well as on a broad review of the existing literature.

The solutions available to monitor e-commerce and to mine data from websites, social media or marketplaces, are numerous. The goal of this report is not to propose an exhaustive presentation of all the solutions that currently exist, but rather to showcase those that our consultations and literature review enabled us to identify.



SOLUTIONS

THE SOLUTIONS

In this section, a series of tools are presented that can help LEAs exploit the potential of data analytic techniques to fight wildlife cybercrime. These techniques can support LEAs in accessing and analysing a large amount of online data, with the potential of greatly reducing the human effort needed to monitor illegal wildlife trade online.

The solutions presented cover three categories:

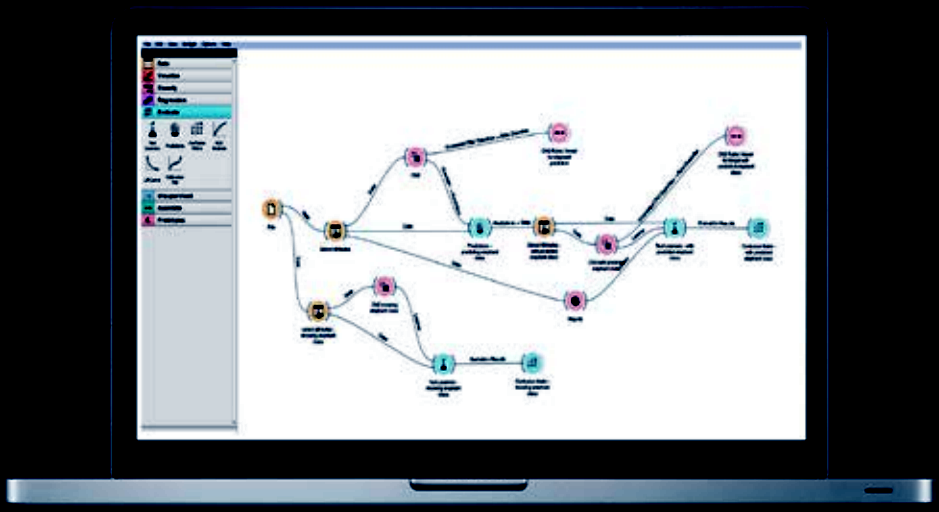
1. **Non-commercial solutions** developed by academia and civil society;
2. **Commercial solutions** which can potentially be tailored to the needs of LEAs;
3. **Projects and tools to develop solutions in-house** with the aim of avoiding externalising sensitive information and support as much as possible the particular needs of various practitioners.

2.1. NON-COMMERCIAL TOOLS

UNIVERSITY OF KENT MODELS

ITRADE

https://peerj.com/articles/cs-10/



KEY STRENGTHS:

This tool is able to **differentiate between legal and illegal goods** with a high level of accuracy.

CHALLENGES

For now, it has **only been tested on limited taxa** (ivory) and marketplaces (eBay).

AVAILABILITY TO LEAS

The **algorithm is freely available through the researchers' paper** (Roberts & Hernandez-Castro, 2015). They have established partnerships with a number of law enforcement agencies including the UK's National Wildlife Crime Unit and Border Force

SOURCE:

The information in this section has been collated from several interviews with David Roberts from the University of Kent and a published article Automatic detection of potentially illegal online sales of elephant ivory via data mining (Roberts & Hernandez-Castro, 2015).

TOOL DESCRIPTION:

A key challenge in wildlife cybercrime lies in determining whether the specimen/product offered for sale is in fact being traded illegally or not (Haysom, 2019). The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and its transposition in the EU through the EU Wildlife Trade Regulations[7], provides several exemptions allowing the trade in protected species; therefore, not all protected specimens found online may be illegal. Verifying the illegality of wildlife trade online is complex as there are very few clear-cut cases where it can be inferred, based on the information provided, that an offer for sale/advertisement definitely involves an illegally sourced specimen.

[7] The text of Regulation (EU) 2016/679 is available here https://eur-lex.europa.eu/search.html?DTN=0338&DTA=1997&qid=1484753427128&DB_TYPE_OF_ACT=regulation&DTS_DOM=EU_LAW&typeOfActStatus=REGULATION&type=advanced&lang=en&SUBDOM_INIT=CONSLEG&DTS_SUBDOM=CONSLEG

Researchers from the University of Kent developed an automated system to detect potentially illegal elephant ivory items for sale on eBay . Two former law enforcement officers, with specific knowledge of elephant ivory identification, manually classified items for sale in the antiques section of eBay UK over an eight-week period. This set the benchmark that the researchers aimed to emulate using data mining. To achieve this, the researchers gave human experts all available information, while they only used metadata[9] for their analysis. This automated system achieved close to 93% accuracy in detecting illegal elephant ivory with less data than the experts, thus proving the potential and replicability of their approach.

Roberts and Hernandez-Castro approached the problem with strictly less information than the human experts used for comparison, which consequently proves the potential of the technique. In this case, metadata included data such as the postage costs associated with the item, its price, whether it was offered through an auction or a buy-it-now, the number of bids received, the number of reviews and feedback about the vendor.

The reported accuracy could be improved with the addition of text mining techniques for analysis of the item description, and by applying image classification (e.g. for the detection of Schreger lines, indicative of elephant ivory). However, this paper was an attempt to find solutions not relying only on images or text descriptions as these could not be employed in other wildlife illegal markets where pictures can be missing or misleading and text absent.

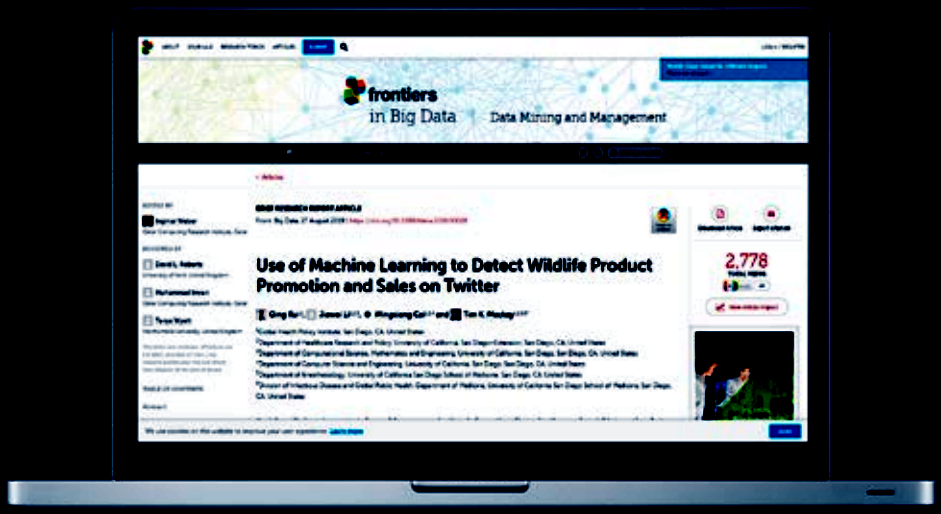
Building on this work, the researchers are developing software called iTrade that automatically identifies likely wildlife cybercrime items posted for sale using eBay as the pilot e-commerce website. Data are obtained via the marketplace's API, and there are plans to expand the scope of the system to include other online marketplaces. The aim is that iTrade would automatically identify likely wildlife cybercrime items posted for sale on any major e-commerce website.

[8] Illegality was based on an assessment by the law enforcement officers as to whether they would take it forward for an investigation e.g. the seller was willing to send the item worldwide or that it was considered not to be antique ivory. Note that in 2009 eBay implemented an internal policy that does not allow the sale of ivory on their website. Therefore, even legitimate ivory cannot be sold on their platform. This has not been considered by the officers when assessing the legality of each item

[9] In this paper metadata refers to all the information associated with an item offered for sale, except the two that are arguably the most informative for the identification of elephant ivory items, namely the item description and item images. Often used in expert investigations, these can be unreliable in some instances according to the authors from the University of Kent. Indeed, on some platforms, such as Instagram or Facebook, there may not be any item description and image recognition is still not capable of detecting ivory in images, although this may change due to the recent advances in machine learning.

UNIVERSITY OF CALIFORNIA SAN DIEGO MODELS

https://www.frontiersin.org/articles/10.3389/fdata.2019.00028/full



KEY STRENGTHS:

This model combines machine learning and NLP to successfully detect illegal content on Twitter without the use of pre-existing—manually built— datasets. This could be key in overcoming the challenge of building critical datasets.

CHALLENGES

This recent model has only been tested on a limited number of taxa (elephant and pangolin), and on one marketplace, Twitter.

AVAILABILITY TO LEAS

The paper referenced above contains the description of the algorithms used. Moreover, the authors are keen on collaborating with LEAs in order to improve their model.

SOURCE:

The information in this section has been collated from a published article Use of machine learning to Detect Wildlife Product Promotion and Sales on Twitter (Xu, *et al*., 2019).

TOOL DESCRIPTION:

In this study, the researchers used the Twitter public API to access Twitter messages in order to detect and classify suspicious wildlife trade and sale using a machine learning topic model combined with keyword filtering and manual annotation. They chose two prohibited wildlife animals and related products: elephant ivory[10] and pangolin, and collected tweets containing keywords and known code words related to these species. In total, they collected 138,357 tweets filtered for these keywords over a 14-day period and were able to identify 53 tweets from 38 unique users that they suspect promoted the sale of ivory products. The study is divided into four phases including: (1) manual search; (2) data collection; (3) data processing; and (4) data analysis.

[10] Please note that some ivory can be traded legally (e.g. antiques) in several countries under certain conditions. For more information please refer to the guidance document produced by the European Commission — EU regime governing intra-EU trade and re-export of ivory (C/2017/3106). Available here https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017XC0517%2801%29

For both the ivory and pangolin datasets, they analysed tweets using the biterm topic model (BTM), which is a machine learning model that uses NLP to categorise short forms of text in a given number of groups (topics) by analysing the correlations between words and topics. They used BTM as a topic clustering method to categorise similar text into related topic clusters.

In machine learning and NLP, a topic model is a type of statistical model for discovering the abstract "subjects" that appear in a collection of texts. Topic modelling is a frequently used text mining tool for discovering semantic structures hidden in a body of text. Intuitively, since a document is about a particular subject, it can be expected that particular words will appear more or less frequently in the document. In a text about elephants, it would not be too surprising to see words like "tusk" or "ivory" appear often. A document generally concerns, explicitly or otherwise, several subjects in different proportions; for example, in a document 20% about tigers and 80% about elephants, there would probably be about four times more words related to elephants than tigers. The topics produced by the topic modelling techniques are groups of similar words. A topic model captures this intuition in a mathematical framework, which makes it possible to examine a set of documents and to discover, based on statistics of the words contained in each, what the subjects may be and what is the balance of the subjects in each document. This model can be used to find short texts (e.g. product descriptions, commentaries etc.) within a large number of other non-related texts.

The results of several studies show that machine learning combined with supplementary analysis approaches such as those utilised in this study have the potential to detect illegal content without the use of pre-existing—manually built—training datasets. If developed further, these approaches can help law enforcement officials, technology companies, and conservation groups, to overcome one of the greatest challenges in developing automated solutions to expedite the process of identifying illegal online sales, which is to build critical datasets.

WILDTRADE

EU-FUNDED PROJECT

KEY STRENGTHS:

This project proposes a comprehensive approach, combining computer vision and natural language processing methods to analyse concomitantly the text and in the visual content of digital platforms.

CHALLENGES

The tools under this project are still in their development phase and have not been tested in real conditions yet.

AVAILABILITY TO LEAS

These solutions will be tailor-made for law enforcement purposes and made available to officers to support them in their endeavours.

SOURCE:

The following information has been collated from consultation with the project co-ordinator, Enrico Di Minin from the University of Helsinki.

TOOL DESCRIPTION:

This group of researchers, co-ordinated by Enrico Di Minin, use computer vision methods for automatic identification, counting, and description of species and wildlife products in images and videos collected from digital platforms, such as social media or online news (Di Minin, *et al*., 2018).

They also use methods from NLP, which refers to the ability of a computer to understand the meaning of human language to extract information about species and wildlife products in multiple languages from digital text content. For example, the University of Helsinki recently used NLP methods to assess sentiment for iconic species from social media and online news automatically (Fink, *et al*., 2019). Combining computer vision and natural language processing methods is an important frontier topic in automatic content analysis with clear potential applications for investigating wildlife cybercrime. The relevant information on illegal wildlife trade, for instance, can be contained in the text or in the visual content of digital platforms, or in both, therefore requiring both visual and textual content analysis.

At the time of writing, the tools and methodologies embodied in the WILDTRADE project are not yet fully developed. These will be publicised and presented in the course of 2020. These solutions will be tailor-made for law enforcement purposes and made available to officers to support them in their endeavours.

WILDEYE

KEY STRENGTHS:

It collects data from a large number of online retail and social media sites to identify illegal species being offered for sale, based on a series of keywords.

CHALLENGES

The final product is not yet available.

AVAILABILITY TO LEAS

This tool has been designed to support the LEAs. The access will be granted to them for free.

SOURCE:

The following information has been collated following consultation with the researcher Jennifer Jacquet from the New York University (NYU).

TOOL DESCRIPTION:

WildEye is a tool that collects data from online retail and social media sites to identify illegal species being offered for sale. It is a secure site and backend algorithm developed by NYU and Indiana University (IU) that scans internet sites to detect trade in wildlife products. The researchers are confident that it can identify illicit products more accurately and at a greater scale than current manual search efforts, improving search effectiveness and saving valuable time and resources.

The tool has two broad sections: 1) a web-based front-end user interface and 2) a back-end data analysis model. The front-end user interface runs on any web browser and can be used to view the ads that have been classified by the WildEye tool as suspicious. The front-end interface is periodically updated with the latest findings. Users can browse, view and search for illegal items identified by the tool. The back-end section of the tool has three sequential phases:

1. Collecting the data from a predetermined set of sites using a list of relevant keywords;
2. Analysing the data to identify potentially illegal ads and classifying them;
3. Pushing the results onto the front end.

The site is currently available in English and scrapes about 100 sites all over the world for CITES Appendix I-listed species. At the time of writing, the team is working on improvements to the user interface as well as the back-end model.

THE DYNAMIC
DATA DISCOVERY ENGINE

https://globalinitiative.net/wp-content/uploads/2019/01/TGIATOC-DetectingOnlineMarkets-Web-1.pdf

KEY STRENGTHS:

This is a methodology that aims to build as comprehensive a picture as possible of how, where and when protected species are transacted over the internet. It consists of a large-scale analysis of how wildlife trade is discussed on the internet. This process also sets out to find keywords and key phrases that best identify wildlife cybercrime activity.

CHALLENGES

This tool cannot yet gather only URLs that are pertinent. In all tested case studies, a substantial amount of irrelevant results within the dataset has remainedirrelevant results within the dataset has remainedmarketplace, Twitter.

AVAILABILITY TO LEAS

The methodology is fully described in the paper referenced above

SOURCE:

The information in this section has been collated from the published report Detecting Online Environmental Crime Markets (Carl, *et al*., 2019).

TOOL DESCRIPTION:

The Global Initiative Against Transnational Organized Crime has composed and trialled a new technical methodology, the Dynamic Data Discovery Engine (DDDE). This is a process that aims to build as comprehensive a picture as possible of how, where and when plants and animals listed in CITES, or commodities containing them, are transacted over the internet. It also aims to identify the broader conversations related to these species and products. By striving to identify, study and understand the discussion forums dedicated to wildlife trafficking on the internet, this tool is the only known solution addressing the socialisation process inherent to wildlife trafficking that creates communities and subcultures where this trade is normalised, routine and unchallenged.

DDDE was developed with the aim of building upon qualitative research to produce larger, more comprehensive datasets of similar activity taking place. It is a six-stage process, that combines a number of different technologies and techniques developed for web-scraping, machine learning and data visualisation into one single and repeatable process. It attempts to allow an analyst to expand a potentially small, pre-existing series of examples of online sources related to the transaction of CITES-listed species into a larger body of examples.

The first stage involves building an initial set of texts (forums discussions, commentaries, items descriptions, etc) that concerns wildlife cybercrime. The second uses feature or phrase extraction to discover the phrases that characterise the original texts. The idea here is to identify what makes these texts undeniably linked to illegal wildlife trade (e.g. the use of certain keywords, key phrasing of sentences, codes, acronyms, prices, etc) in order to be able to spot the same narrative elsewhere. Thirdly, these features are combined to create search phrases; web searches are performed; and the results of these searches are collected. Next, a series of language-based forms of analysis are used to separate the relevant from the irrelevant results in these search queries. Then more qualitative forms of analyses are manually performed on the relevant outputs to produce a more condensed, detailed appraisal of the data collected. In the final, sixth stage, the data are visualised. The process is recursive, working "in cascade" where the outputs of one iteration are used as inputs for the next one, allowing an analyst to return to any stage in the pipeline and update, add or filter the data further. This cyclical nature of the DDDE is intended to allow it to evolve and improve over time.

The identification of online environmental crime seems to be well suited to the language-driven way that the DDDE works. The overall process sets out to find keywords and key phrases that best identify wildlife cybercrime activity, and a key subset of these are code words, or a body of language intended to mask this purpose. Indeed, although this kind of language is intended to camouflage the nature of the illicit activity, this project found that, conversely, it could be highly indicative of it.

However, a key weakness of the DDDE process is that it cannot currently gather only Uniform Resource Locators (URL, more commonly named web addresses) that are relevant. In all tested case studies, a substantial amount of "noise" within the dataset has remained. Therefore, despite the large absolute number of relevant URLs identified, they constitute a relative minority within the final overall datasets. Until techniques to remove this noise become more reliable and effective, it would be difficult for a human analyst to be able to act on the data that the DDDE produces. To become actionable or operationally useful, it is important that the process becomes better at guiding human end-users in sorting and prioritising the information that it produces.

CHIMPFACE

https://conservationx.com/project/id/8[11]

KEY STRENGTHS:

This recognition software has been able to successfully recognise species as well as individual Chimpanzees based on photos published on the internet.

CHALLENGES

This project is still very early in its development. Also, it would require enormous datasets in order to be scaled-up to other taxa.

AVAILABILITY TO LEAS

The ultimate goal is to support for free the law enforcement community leading to prosecutions of traffickers.

TOOL DESCRIPTION:

ChimpFace is facial recognition software that has been developed to recognise species as well as individual Chimpanzees when photos published on the internet contain Chimpanzees.

This tool focuses on great apes. However, it could potentially be replicated to other species. and is a promising example of the potential of computer vision methods for visual content analysis.

ChimpFace developers believe that live great apes frequently fall victim to Internet-enabled trafficking. Researchers have identified hundreds of social media accounts displaying illegally traded great apes and have manually searched millions of social media photographs for evidence of trafficking.

ChimpFace aims to use image analysis techniques (facial recognition) to identify photos on social media and e-commerce websites that are likely linked to trafficking-related activities. Suspicious posts will be automatically flagged as possible trafficking and sent to a team of conservation experts for review. The initial phase of the project focused on Chimpanzees because of their importance to conservation, the frequency at which they

[11] The developers of this tool are Dr. Alexandra Russo and Dr. Colin McCormick, affiliated to Conservation X Labs

appear in sales online, and the availability of large amounts of imagery. The project team leveraged both publicly available image datasets and images collected from Chimpanzee conservation and research organisations to build a binary Chimpanzee image classifier and then deploy it to monitor social media posts. Depending on the number of posts that are flagged by the image analysis method, they may also add text-analysis to their method to down-select posts based on likely trafficking-related words or phrases (e.g. "chimp for sale").

This project, developed in the USA, is still very early in its development. However, their ultimate goal is to support the law enforcement community leading to prosecutions of traffickers.

2.2. COMMERCIAL TOOLS

The number of private companies providing commercial services for web-scraping, data mining or machine learning is growing. Such services are largely tailored to commercial purposes and used by companies as a competitive intelligence tool. Yet, since such tools can potentially be adapted and used to monitor wildlife trafficking online, engaging an external company for this purpose can be an option for LEAs. This section will be dedicated to sharing the experiences of wildlife cybercrime experts, both from LEAs and NGOs, rather than listing existing business solutions. This report is not intended to promote private company services ahead of others, or to provide a market study and a comparison of the actors that compose it. Furthermore only a few of the tools for online monitoring, of which we have knowledge, have in fact been tested in the field of wildlife cybercrime. Hence, this section will be dedicated to sharing the experiences of wildlife cybercrime experts, both from LEAs and NGOs, rather than listing existing business solutions.

Regarding the solutions mentioned, it is noted that they have not been selected because they are specifically adapted for the fight against wildlife trafficking online. These are generic tools, which offer the opportunity, and the flexibility, to be adapted to many different topics, including wildlife cybercrime. Their inclusion in this report, therefore, is rather due to their technical expertise in data mining and management than their ability to work on wildlife cybercrime.

The results on the use of commercial tools for monitoring wildlife trafficking online are relatively limited, for three main reasons put forward by the interviewees.

First, working with such companies when representing a public body implies the respect of many procurement regulations, including some company certification rules. Then, the price is usually higher than for collaborations with the academic world. And finally, flexibility, responsiveness and working conditions are described as better when collaborating with researchers.

Netherlands Food and Consumer Product Safety Authority
Ministry of Agriculture, Nature and Food Quality

EXPERIENCE FROM THE
NETHERLANDS FOOD AND CONSUMER PRODUCT SAFETY AUTHORITY

The Netherlands Food and Consumer Product Safety Authority (De Nederlandse Voedsel- en Warenautoriteit, NVWA) is collaborating with a small company delivering tailor-made web-scraping and monitoring services[12]. Working with a commercial supplier allows NVWA not to concern itself with employing IT professionals, working through technical details and administering the system. It provides the possibility to experiment with web-scraping and how to process the data output. Analysing the data output is their biggest challenge.

NVWA's CITES experts and their contractor have been developing a tailor-made monitoring tool for tracking advertisements and sales of exotic animals and plants[13]. The monitoring tool has given them insights into different online domains (Dutch) where individuals are trying to sell these plants and animals. The problems persist within the context of online enforcement and validating the data. There are a lot of useless data for which they are yet to find efficient ways to analyse automatically.

In order to tackle these bottlenecks, The NVWA is going to work together with the Wageningen University & Research (WUR), that has already shown promising results by developing solutions based on a Bayesian Network[14]. In 2020, NVWA are planning to further develop and consolidate their monitoring tool.

ifaw

EXPERIENCE FROM
INTERNATIONAL FUND FOR ANIMAL WELFARE (IFAW)

IFAW has collaborated with a company providing anti-cybercrime services in France, called Webdrone, to help gather data for their Disrupt: Wildlife Cybercrime report, published in 2018 (IFAW, 2018)[15].

Collaboration with this company specializing in data mining encountered several challenges. Firstly, these cybercrime experts were not experts in the field of CITES and therefore did not have any previous knowledge on CITES related topics. Consequently, it took time to train the people in charge of the scraping operations to implement the methodology developed by IFAW.
Secondly, the results provided, contained a number of inaccuracies despite the trainings and the key words provided by IFAW. There was a lot of information that needed to be sifted through to correct mistakes and inaccuracies. It required quite an extensive number of hours of cleaning work by IFAW teams in France. Ultimately, the use of this tool did not reduce the

[12] At the time of writing, the NVWA was in the process of developing its own software, which could be available in 2020, and potentially shared with their European counterparts.

[13] More information is available here: https://www.nvwa.nl/onderwerpen/invasieve-exoten/unielijst-invasieve-exoten.

[14] A Bayesian network is a popular probabilistic model that represents a set of data and their conditional dependencies. In the context of wildlife, the presence or variation of certain data could lead to inferring, from a probabilistic perspective, the presence of illegal trade. For example, a sudden change of price for specific species might suggest disruption from an illegal market. For more information on this, please refer to https://en.wikipedia.org/wiki/Bayesian_network

[15] The report is available here https://www.ifaw.org/news/disrupt-wildlife-cybercrime-report

workload for IFAW, but rather changed its nature. According to the IFAW officer in charge of this project, the complexity of the subject, particularly the legislative context, is perhaps one of the main reasons for the relative lack of precision of the work provided by this data mining tool. However, it must be noted that the grey area between legal and illegal is always a challenge for any type of online research whether carried out by data mining tools (automatic or semi-automatic) or purely manual methods.

Ultimately, IFAW team remains convinced that the result of this experiment was encouraging, if not totally satisfactory, and that more resources (both financial and human) should be invested in the development of such automated tools.

2.3. IN-HOUSE DEVELOPMENT OF SOLUTIONS

Developing software in-house can guarantee a more tailored approach to the specific needs of the LEAs. Such approach avoids the sharing of sensitive information or key processing methodology with external third parties. In some cases, it can reduce procurement costs as well as administrative overheads. LEAs in the EU interested in developing in-house solutions must make sure to comply with the GDPR requirements and Police-Justice Directive described in section 1.1.1.

2.3.1. TITANIUM PROJECT

INTERPOL's Cyberspace and New Technologies Lab works to develop innovative policing tools for its 194 member countries. This laboratory is a consortium member of the EU Horizon 2020 TITANIUM project (https://www.titanium-project.eu/).

TITANIUM researches, develops and validates novel data-driven techniques and solutions designed to support LEAs investigating criminal or terrorist activities involving virtual currencies and/or underground markets on the darknet.

The result of TITANIUM is a set of services for data mining and forensic tools, which operate within a privacy and data protection environment that is configurable to local legal requirements, and can be used by investigators for:

- ✓ Monitoring trends in virtual currencies and darknet market ecosystems;
- ✓ Analysing transactions across different virtual currency ledgers;
- ✓ Generating court-proof evidence reports based on reproducible and legally compliant analytical procedures.

Two of the tools developed under the TITANIUM project are now available to the law enforcement officers.

DARK WEB MONITOR

Dark Web Monitor provides data for different use cases, leading to strategic insights, deep understanding of tactics and operational support to identify actors based on mistakes made in the past.

- ✓ Dark Web Monitor gathers information about online activities on crime areas as drugs, weapons, cybercrimes and counterfeiting;
- ✓ Hidden services (>200k), Forum posts (>30M), Usernames (>500k) and many entities as cryptocurrency addresses, email addresses, etc;
- ✓ Search, explore and API functionalities for full integration with LEAs infrastructure.

Accessing Dark Web Monitor requires a subscription[16]

[16] More details are available on "https://dws.pm/monitor/".

GRAPHSENSE

GraphSense is a cryptocurrency analytics platform with an emphasis on full data sovereignty, algorithmic transparency, and scalability. GraphSense is open source and free. It provides a dashboard for interactive investigations and, more importantly, full data control for executing advanced analytics tasks[17].

2.3.2. OPEN SOURCE SUPPORTING TOOLS

There are many open source tools that can help LEAs to develop their own in-house tailored solutions.

Here is a non-exhaustive list of tools that can retrieve, parse and analyse data from the internet to shed light on the existing possibilities. It is important to note that the tools below are not dedicated to wildlife cybercrime but have the potential to be adapted to this subject.

Scrapy (https://scrapy.org/): is a web-crawling framework. Originally designed for web scraping, it can also be used to extract data using APIs. This tool is relatively simple to use as it does not require any advanced knowledge in coding language (Python, in this case). The website proposes a course, free of charge, to get to know the tool.

Hyphe (http://hyphe.medialab.sciences-po.fr/) is free software for web crawling and corpus curation. It was originally designed by the Institute for Political Sciences in France to provide researchers and students with a tool for creating and cleaning a set of texts based on a research-oriented crawler. Rather than dealing with whole "websites", Hyphe handles "web entities", which can be defined just as simple pages, as subdomains, or a combination of sites. In the case of wildlife cybercrime, it could be a list of internet sites, identified as hosting relevant illegal activities. The pages residing under these web entities are then automatically crawled, in order to collect the outgoing links and the text contents.

The GATE project (https://gate.ac.uk/) offers open-source text analysis tools. GATE is an infrastructure for developing and deploying software components that process human language. It has been developed by the University of Sheffield, nearly 15 years ago, to address comprehensively all types of computational tasks involving human language. GATE can a priori tackle text analysis of all shapes and sizes. This tool, open source, is supposedly able to solve any form of problem relating to the analysis of text.

2.4. SUPPLEMENTAL SOLUTIONS

None of the solutions presented above provides the silver bullet for the fight against wildlife cybercrime. Despite significant progress in this area, and multiplication of initiatives both within academia and the private sector, these solutions alone cannot solve this problem. Other solutions can be used, complementarily, to the fight against wildlife cybercrime.

This chapter aims to highlight these other technological solutions. They fall in the following categories:

1. Communication and collaboration tools;
2. Identification tools;
3. Investigation tools;
4. Tools tackling the socialisation processes;
5. Other tools

[17] For more information on Graphsense, contact Mr. Bernhard Hashofler on "https://graphsense.info".

2.4.1. COMMUNICATION AND COLLABORATION TOOLS

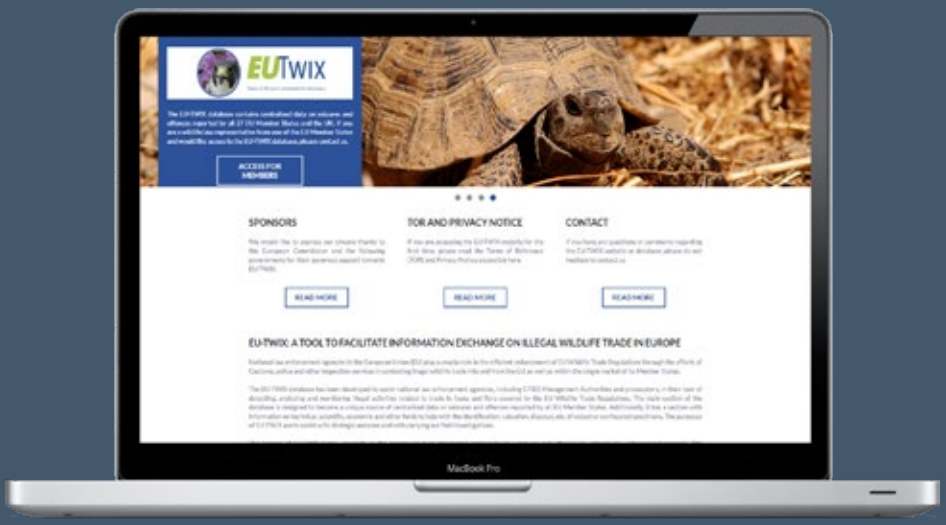
Communication tools between LEAs could play a two-fold role: first in managing, sharing and processing large volumes of data, often held by different agencies, and, second, in improving the ability to report wildlife cybercrime.

Some tools, mostly institutional, have been developed to facilitate communication between the agencies involved in the detection and investigation of wildlife crime. The concerted use of these tools is key to realising the full potential within data held by LEAs. Gathering data to enter into automated algorithms is one of the most challenging aspect of fighting wildlife cybercrime; a very high amount of data is needed to input into the tools, while this high amount of data is currently missing. Communication tools for LEAs such as EU-TWIX and CENcomm (ENVIRONET) can play an important role in assembling and organising the data already held by the LEAs. Furthermore, communication tools between LEAs can help improve the exchange information to understand better the phenomenon of wildlife cybercrime and the modus operandi of traffickers.

TRADE IN WILDLIFE INFORMATION EXCHANGE

EU-TWIX

https://www.eu-twix.org/



KEY STRENGTHS:

It enables near real-time exchange of information between enforcement and management officials. Types of information shared include seizure details such as modus operandi, stolen specimen alerts and assistance requests for species identification.

CHALLENGES

This tool does not allow nominal information to be shared.

AVAILABILITY TO LEAS

This tool was developed for the exclusive use of wildlife law enforcement and management officials. The access is free of charge.

SOURCE

The information in this section has been collated from the referenced webpage and consultation with the EU-TWIX project manager.

SOURCE

EU-TWIX (https://www.eu-twix.org/) is an enforcement support system in Europe developed for the exclusive use of wildlife law enforcement and management officials. At the time of writing, the system connects a network of over 1200 officials from 39 European countries.

The EU-TWIX mailing list enables near real-time exchange of information between enforcement and management officials. Types of information shared include seizure details such as modus operandi, stolen specimen alerts and assistance requests for species identification. The EU-TWIX mailing list brings cases to the attention of agencies in countries which unknowingly played a role in trade and has triggered numerous investigations to date. The tool is recognised globally as an exemplary communication channel and has been replicated in other regions; it is used daily by enforcement officials to communicate efficiently and effectively with their colleagues across Europe.

The EU-TWIX website holds a database which centralises information on wildlife trade seizures submitted by enforcement agencies including the police, customs, and environmental inspection services. Access to the website is exclusively granted to designated officials who are provided with access codes. The EU-TWIX website also holds information on laboratories, rescue centres and wildlife experts, as well as prices of specimens in trade. Other types of resources available via the website include identification tools and training materials (including on wildlife cybercrime).

(ENVIRONET)
CENCOMM

http://www.wcoomd.org/fr/topics/enforcement-and-compliance/instruments-and-tools/cen-suite/cencomm.aspx

KEY STRENGTHS:

It allows Customs officers and other authorized users to exchange real-time information during an operation. It also provides a means to share information in a standardized format

AVAILABILITY TO LEAS

This tool was developed for the exclusive use of wildlife law enforcement and management officials. The access is free of charge.

SOURCE

The in this section information has been collated from http://www.wcoomd.org/fr/topics/enforcement-and-compliance/instruments-and-tools/cen-suite/cencomm.aspx

TOOL DESCRIPTION:

The Customs Enforcement Network Communication Platform (CENcomm) is a secure online platform allowing Customs officers and other authorized users to exchange real-time information during an operation. It also provides a means to share information in a standardized format. CENcomm has several programme specific applications; one of them is ENVIRONET.

ENVIRONET is a tool for information exchange and cooperation in the area of environmental issues specifically, modelled after EU-TWIX. As one of the CENcomm applications, ENVIRONET is accessible only to a closed user group. Information transmitted via the tool is encrypted. The tool enables officials across the world to exchange information relating to wildlife crime (amongst others), such as seizures alerts and identification tools.

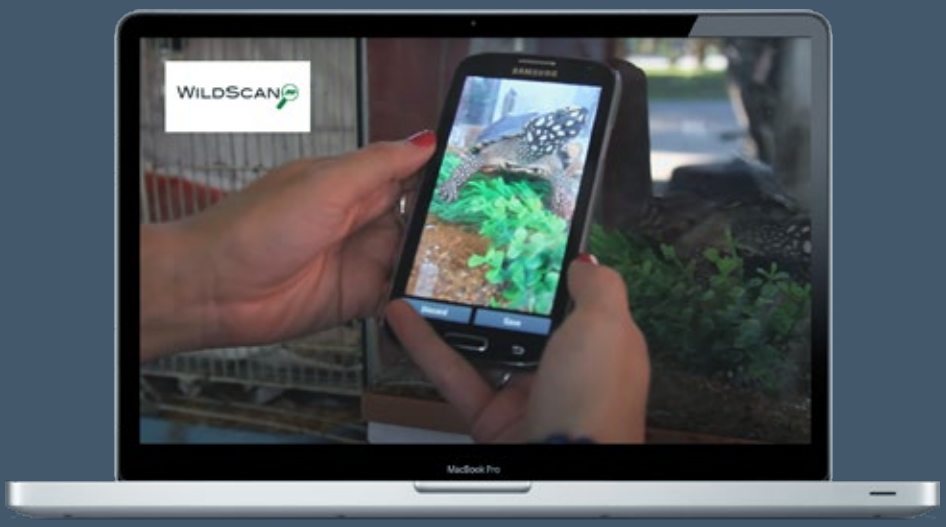
2.4.2. IDENTIFICATION TOOLS

Several initiatives today aim to support the identification of protected species, including through applications for phones, in order to fight against wildlife trafficking. Some offer the ability to scan directly with the camera of a mobile phone, while others, based on a series of questions asked to the observer can deduce species.

MOBILE APP

WILDSCAN

https://wildscan.apk.cafe



KEY STRENGTHS:

The application makes it possible for everyone, including citizens, to easily identify and report wildlife trafficking and support ongoing law enforcement efforts to counter wildlife crime.

CHALLENGES

Not targeting the EU.

AVAILABILITY TO LEAS

Accessible to everyone, free of charge.

SOURCE

The information in this section has been collated from the application reference page https://wildscan.apk.cafe/, and a report published by the World Bank (The World Bank, 2018).

TOOL DESCRIPTION:

WildScan is a mobile phone application that identifies physically seen live specimens of endangered species. It was designed with the aim of helping LEAs combat wildlife trafficking. WildScan contains photos and information about more than 280 endangered species and illegal wildlife products commonly trafficked into and throughout Southeast Asia. The App aims to help identify the animals and facilitate a rapid response for their rescue.

The application makes it possible for everyone, including citizens, to report wildlife trafficking and support ongoing law enforcement efforts to counter wildlife crime. WildScan was produced through a collaborative partnership between academics, law enforcement, scientists and other wildlife specialists under the USAID-funded Asia's Regional Response to Endangered Species Trafficking (ARREST) Program.

The application allows users to input several types of information such as the colour and size of the animal to quickly identify the species. It also includes essential animal care instructions and a simple reporting function. Although this tool does not concern Europe, it nevertheless remains interesting as a source of inspiration for the development of a similar application, concerning the most traded species in Europe on the internet.

MOBILE APP

WILDLIFE ALERT

https://apps.wcswildlifetrade.org

KEY STRENGTHS:

It helps users to assess the authenticity and the source of wildlife items. If an item is thought to be from a protected species, the user is provided with instructions on how to proceed.

CHALLENGES

The app works on a limited number of species and requires some training.

AVAILABILITY TO LEAS

Accessible to everyone, free of charge.

SOURCE

The information in this section has been collated from the application reference page https://apps.wcswildlifetrade.org/, and a report published by the World Bank (The World Bank, 2018).

TOOL DESCRIPTION:

The Wildlife Conservation Society's Wildlife Alert App poses a series of questions regarding wildlife products which users have come across. These questions help the user assess the authenticity and the source of the item in question. If the item is thought to be from a protected species, the user is provided with instructions on how to proceed. Wildlife Alert provides users with quick access to useful information on hundreds of protected wildlife species. It is seen as broadly reliable and require little training.

2.4.3. INVESTIGATION TOOLS

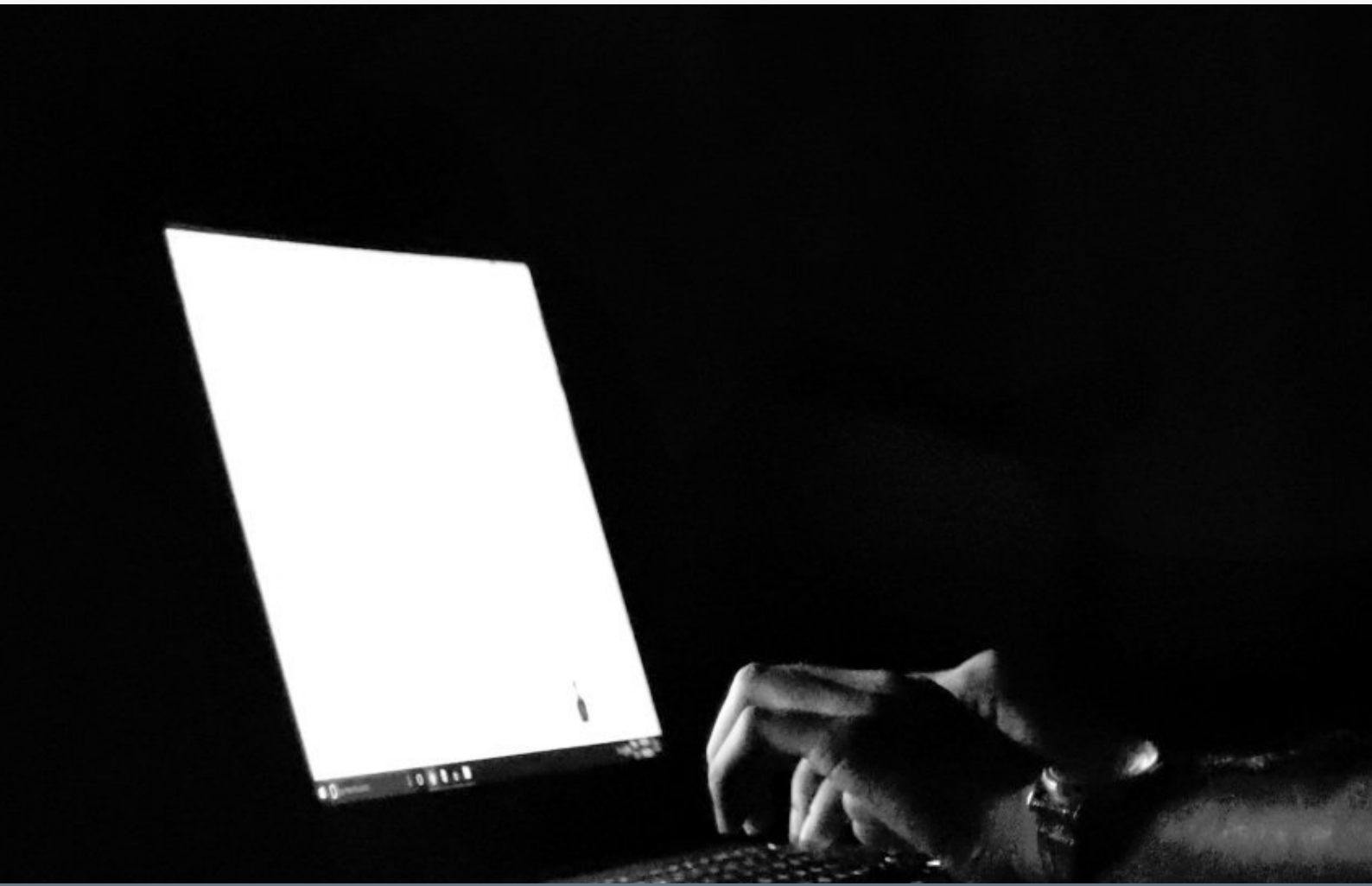
To investigate internet-related crime, law enforcement needs to be able to search computer networks; intercept, collect and keep communications data; and have the power to seize assets, such as computers and phones. Internet-based trafficking adds to existing challenges around wildlife crime enforcement and compels enforcement officials to operate in a cross-jurisdictional, virtual space that they are largely unprepared to manage (Wingard & Pascual, 2018). The solutions presented below go one step further than the sole online monitoring of wildlife trafficking online and proposes solutions to investigate within the vast cyberspace. While web scrapers help to detect crime, the tools presented in this section help to conduct the investigation.

Some experts claim that online investigation practices are out of step with the amount of data that needs to be captured and processed (Haysom, 2019). Given the vastness of cyberspace, investigation techniques that have grown from historic models in the offline world are outpaced; the amount of information now potentially available, instead of being an asset to investigators represent an obstacle.

Several data processing tools inspired by research on big data analysis techniques are aiming to simplify the task of investigators.

Aware that online investigation tools are to a large extent known to investigators, it is still deemed useful to highlight a few of the tools available which can help investigators in their daily work, in case such tools were still unknown or under-exploited in the context of wildlife cybercrime investigations. It is important to specify that these investigative tools have not been designed to be applied on a particular topic, they can be applied to a variety of issues.

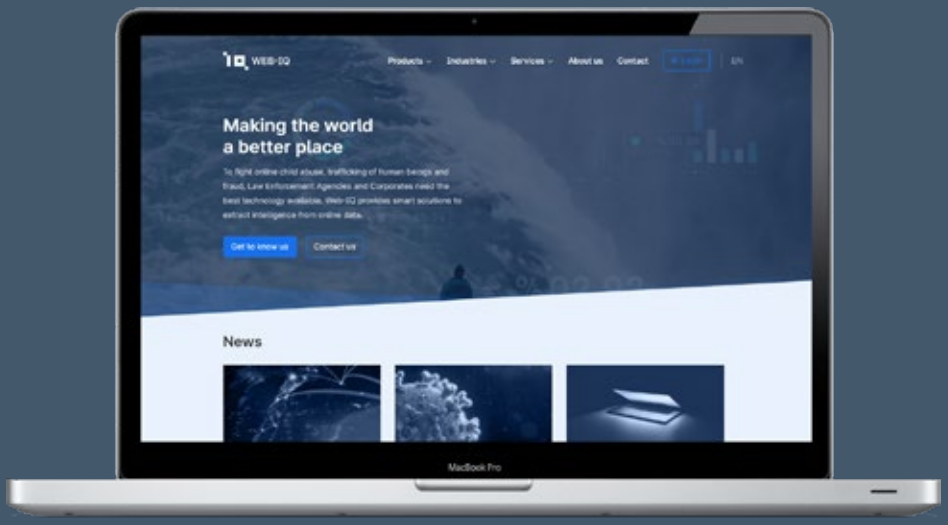
The tools listed below are only a small sample of the options available.



MOBILE APP

WEB-IQ

https://www.web-iq.eu



KEY STRENGTHS:

This tool offers reliable and customised open source intelligence (OSINT) investigation support.

CHALLENGES

Never tested on wildlife cybercrime.

AVAILABILITY TO LEAS

Web-IQ is a commercial tool.

SOURCE

The information in this section has been collated from the commercial webpage https://www.web-iq.eu.

TOOL DESCRIPTION:

Web-IQ was founded in 2011 in reaction to encountering images of child sexual abuse during a routine crawl of the Web. Web-IQ is working today with a plethora of law enforcement organisations, by offering open source intelligence (OSINT) investigation support.

By combining their expertise and their own technology, they can provide specialised projects to deliver customised on-site solutions for web intelligence.

The models developed and used for tackling child sexual abuse, could potentially be adapted to address wildlife cybercrime.

ANITA PROJECT

https://www.anita-project.eu

KEY STRENGTHS:

This European project is working on delivering a set of advanced tools and techniques to efficiently address online illegal trafficking.

CHALLENGES

Not yet targeting wildlife cybercrime.

AVAILABILITY TO LEAS

The ambition is to support the law enforcement community. The information available so far does not allow assessing whether the access to these tools will be free of charge or not.

SOURCE

The information in this section has been collated from the project webpage https://www.anita-project.eu.

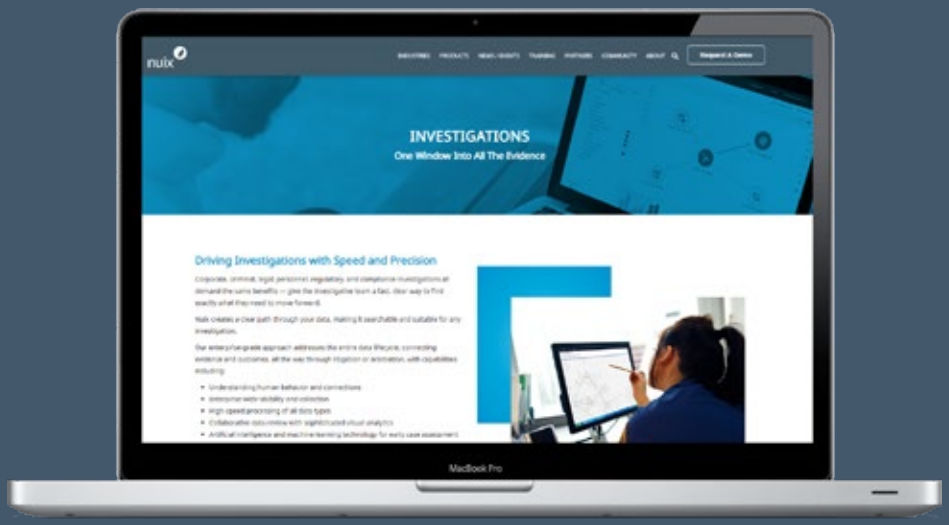
TOOL DESCRIPTION:

ANITA is an EU-funded project that aims at improving investigation capabilities of LEAs by delivering a set of tools and techniques to efficiently address online illegal trafficking.

Their method is comprising knowledge modelling and reasoning services; discovery and monitoring of new and existing online marketplaces; resolving criminal identities in social networks and web as well as identification of authors and web contents; unmasking of fake information, disinformation and camouflage of the real nature of information (e.g. code words); insights on criminal groups relevant and related to trafficking of illegal products; discovery and understanding of trends and behavioural patterns; revealing, tracking, and monitoring of payments and transactions in crypto-currency networks; interoperability with available relevant investigation systems already in place and operation at and for LEAs. This is done with the hope of supporting the LEAs in more effective investigation activities by using online contents.

The project is currently in its pilot phase, focusing on counterfeit/falsified medicines, new psychoactive substances, drugs, and weapons. The methods used and developed within the project could inspire project developers and LEAs for their extension to the field of online wildlife trafficking.

NUIX
AND OTHER SOLUTIONS
https://www.nuix.com/solutions/investigations



KEY STRENGTHS:
Nuix provides an easy-to-use and collaborative digital forensics tool.

CHALLENGES
Nuix has only been tested on wildlife crime on a limited number of occasions.

AVAILABILITY TO LEAS
Nuix is a commercial tool.

SOURCE
The information in this section has been collated from interviews with David L. Robert from the University of Kent, and Stuart Clarke, Chief Technical Officer at Nuix.

TOOL DESCRIPTION:
The Nuix Lab (https://www.nuix.com/solutions/investigations) works on large data-sized investigations. It targets digital forensic looking to build or upgrade a dedicated digital forensics facility.

The core technologies of the Nuix Lab, Nuix Workstation and Nuix Investigate, give digital forensic technicians and case investigators different lenses into the same case data. Investigators benefit from an easy-to-use browser experience where they can collaborate on the same data at the same time, creating efficiency and helping them share insights.

Nuix is a data analytics and investigation software company that enables law enforcement and government agencies, and corporations find truth in their data. The Nuix Investigation lab is a proven collaborative investigation solution in the fight against all forms of organized crime.

The Nuix Investigation Lab helps investigators to work on gigabyte to terabyte-sized investigations. It is meant to support local or small regional forensic labs with limited capacity to tackle the expanding volume, variety, and complexity of digital evidence. The Nuix Lab can help putting together evidence that allow officers to understand, for instance, who is talking to whom, when and how often, and to find seemingly hidden connections using investigative analytics across all case data. Nuix can work alongside other specialist tools and helps investigators pull together the data they produce in a single investigative window into the evidence.

Nuix has already been tested on wildlife trade-related issues, such as the rhino horn trafficking.

Other solutions exist, similar to NUIX, such as I2.

i2 is a suite of software owned by IBM (https://www.ibm.com/security/intelligence-analysis/i2) that facilitates the collection of data, the sharing of information, the discovery of networks and patterns, and the production of intelligence. Originally designed for law enforcement and military intelligence agencies, the software includes a range of data management applications and analytical tools. These tools provide a secure, multi-user environment, facilitating the analysis of large volumes of data and providing innovative features such as connected network visualisations, social network analysis and geospatial or temporal views of information. Today, IBM i2 is widely used around the globe, with clients that include government and industry leaders in key areas such as enforcement, defence, national security, banking, retail, insurance, healthcare and life sciences.

2.4.4. TACKLING SOCIAL PROCESSES

Conducting illicit trade requires a customer base that is socialized into the norms, practices and practicalities of the trade. Some claim that the Internet is a powerful vehicle for creating new demand by socialising people into illegal trade. Research has shown that online platforms allow global demand opportunities to be opened up, turning what might have previously been a small interest group into a large global community of potential buyers (Wingard & Pascual, 2018). This process, facilitated by online platforms, is reinforced by a certain sense of impunity, inherent to wildlife cybercrime due to the lack of enforcement for such crime. This sense of impunity has spread among those on the demand-side of illegal wildlife trafficking, mainly related to the complexity of the regulations associated with this trade and to the fact that, in the context of the pet trade, most buyers consider themselves nature lovers and therefore unable to commit an environmental crime (Haysom, 2019). For example, discussions among enthusiasts who share admiration for rare live specimens, if left alone, can lead to spiralling demand, as enthusiasts seek to attain the same prestige and status as their online contacts by owning and collecting their own rare specimens.

In addition, the trafficking networks that promote the trade and consumption of wildlife are ultimately platform agnostic. Felipe Thomasz explains, in his paper on Illicit Wildlife Markets and the Dark Web that they move to where consumers are, and to where it is the easiest to conduct business, with no loyalty and with negligible investment in using a particular

social-media site, online forum or e-commerce platform (Thomasz, 2018). Thomasz insists that successful interventions in the area of wildlife cybercrime are likely to include the social processes and the underlying consumer psychology, rather than focusing only on platform control and regulation, as that would serve only to displace the location of the network, rather than stop it functioning.

The presence of an "enforcement feeling" is therefore crucial as the lack of visible enforcement online feeds into the normalization of the trade. 'If would-be consumers see a plethora of illegal products being marketed so brazenly, how can they come to view their own purchasing of these goods as a "real" crime?' (Haysom, 2019).
Along with the internet's formidable potential to carry out viral environmentalist campaigns and spread information, several tools have the potential to tackle the social processes active in the context of wildlife cybercrime. Some of the most concerned companies have implemented publicized measures to deter illegal trade, such as the pop-up notifications that appear on Instagram when you try to search for animals threatened by illegal trade (Daly, 2017), or the recent announcement by Tencent that it will introduce a reporting function for users to flag suspected illegal activity (Xinhua, 2018). Both companies are members of the Coalition to End Wildlife Trafficking Online[18] which encourages such measures from online technology companies to ensure that they are part of the solution to end wildlife cybercrime.

Reporting phone Apps can also be considered. Wildlife Witness[19], for example, has been developed by the Taronga Zoo, in 2013, in collaboration with TRAFFIC to be the first global community action smartphone app in the fight against illegal wildlife trade. The Wildlife Witness app allows tourists and locals to easily report illegal wildlife trade by taking a photo, pinning the exact location of an incident and sending these important details to TRAFFIC. The same principles could be applied to the online traffic, where internet user could share screenshots and details about suspicious ads. The Cyber Spotter programme[20] of the Coalition to End Wildlife Trafficking Online is an example of such an initiative, where citizens can report suspicious content they find online directly to the Coalition for removal by companies that are members of the Coalition.

In order to improve the set of measures available to tackle the socialisation aspect, some authors, like Simone Haysom, are calling to study the body of practices built up in dealing with other online crimes – such as child pornography and counterfeit goods – as potential sources of inspiration for wildlife cybercrime (Haysom, 2019).

[18] The Coalition to End Wildlife Trafficking Online, convened by WWF, TRAFFIC and IFAW, brings together the world's biggest e-commerce, technology and social media companies to shut down online marketplaces for wildlife traffickers. More information on the Coalition website at https://www.endwildlifetraffickingonline.org/

[19] The App can be download here - https://apps.apple.com/us/app/wildlife-witness/id738897823

[20] For more information about the Cyber Spotter programme of the Coalition to End Wildlife Trafficking Online visit the following page: https://static1.squarespace.com/static/5b53e9789772ae59ffa267ee/t/5c912f9e9140b78a7bf096b1/1553018782710/Wildlife+Cyber+Spotter+One+Pager+Mar+2019.pdf

THE GLOBAL DATABASE OF EVENTS, LANGUAGE, AND TONE (GDELT) PROJECT

https://www.gdeltproject.org

KEY STRENGTHS:

Monitors print, broadcast, and web news media in about 100 languages from across reportedly every country in the world.

CHALLENGES

Not calibrated to monitor wildlife crime related news.

AVAILABILITY TO LEAS

This tool is accessible to everyone, free of charge.

SOURCE

The information in this section has been collated from the project's webpage https://www.gdeltproject.org.

TOOL DESCRIPTION:

The GDELT project monitors print, broadcast, and web news media in about 100 languages from across reportedly every country in the world to keep continually updated on breaking developments over most of the planet. Through its ability to leverage the world's collective news media, GDELT moves towards a more global perspective on what is happening and how the world is feeling about it. The GDELT Project is an initiative with the ambition to construct a catalogue of human societal-scale behaviour and beliefs across all countries of the world, connecting every person, organization, location, theme, news source, and event across the planet into a single network that would capture what is happening around the world, what its context is, who is involved, and how the world is feeling about it.

The GDELT project could be turned into a powerful tool, if set accordingly, to understand better the scope and nature of wildlife cybercrime, as well as the type of persons involved. Collaboration between wildlife cybercrime practitioners and this initiative could potentially be significantly helpful to understand the relationships between online and offline trade, the routes used, and, how these issues are regarded by the different populations. It could help governments develop effective response strategies.

ALEPH SEARCH ENGINE

KEY STRENGTHS:

Aleph is a search engine that references the dark and the deep web.

CHALLENGES

So far it mainly focused on issues related to drugs, weapons, and terrorism, not yet on wildlife crime.

AVAILABILITY TO LEAS

Aleph is a commercial tool.

SOURCE

The information in this section has been collated from the commercial webpage and was recommended by an independent consultant, interviewed in the context of our research.

TOOL DESCRIPTION:

Aleph is a search engine that references the dark and the deep web; about 2 million sites are indexed for the use of intelligence services and cybersecurity firms, industrial security, cyber-intelligence. This is described as a customizable tool that offers a map of the dark web and could give access to sensitive and strategic data.

https://www.anima.couriers.com/protect-yourself-from-animal-scams/

Today, most of Aleph's customers remain state administrations, although their skills seem to interest more and more the private groups.

At present, this company is mainly focused on issues related to drugs, weapons, and terrorism, but aware of the battle against wildlife cybercrime, one of their representatives mentioned that their tool could be effectively adapted to the problem of trafficking of protected species on the dark web by mapping activities related to wildlife cybercrime on the dark and deep web

ONE-CLICK TOOL

KEY STRENGTHS:

The software can download the details of a suspicious webpage, in a few seconds, in an organised and consolidated fashion, thus saving a considerable amount of time that would otherwise be spent copying and pasting the text and downloading images.

CHALLENGES

This tool has only been tested on eBay, so far.

AVAILABILITY TO LEAS

The ambition of the developer is grant free access for the law enforcement.

SOURCE

The information in this section has been collated from a consultation with the project leader, David L. Roberts, from the University of Kent.

TOOL DESCRIPTION:

The University of Kent have developed a tool to aid investigators and researchers looking at eBay. The aim of the tool is to allow decisions to be recorded, such as whether someone thinks an item is legal or illegal, along with notes to explain their decision-making. The software then downloads the item's details in seconds thus saving a considerable amount of time that would otherwise be spent copying and pasting the text and downloading images.

Before starting, if required, the tool can be set with classifiers based on what an individual wishes to capture based on a question, for example "Is this ivory?" with a classifier of Yes, No or Maybe. If required a conditional classifier can be added, for example if you have said Yes to ivory, you may then want to report that it is Elephant, Hippo, Walrus, etc. or whether it is legal with Yes, No or Maybe. In addition, as images contain a lot of useful information that aid decision-making, you can report the images that aided in this decision by clicking on the image numbers. Finally, there is a comments box for any further comments. Once all the decisions have been made, the item can then be saved. The item is downloaded as a zip folder containing the text (including the decision) in a csv file, all images, the page as screenshot and the page also archived.

Finally, all this information is brought together in a pdf report. Law enforcement officers and others may not require all this information in terms of constructing the decision-making process and therefore selected items

of interest can be downloaded. However, researchers may require this information to create a dataset for machine learning. In effect this speeds up the time by not having to copy and paste data from an item as well as downloading information. In the future, there will be the option to hash the files, which is the digital equivalent of putting the information in an evidence bag, produce reports in the exact format of a police report and then extend the function by allowing basic open source intelligence analysis.

IVORY CROWDSOURCING TOOL

KEY STRENGTHS:

This tool uses crowdsourcing to identify elephant ivory and train computer to recognize ivory..

CHALLENGES

This tool is in its early development and has only been tested on one product (ivory) and one platform (eBay).

AVAILABILITY TO LEAS

The ambition of the developer is grant free access for the law enforcement.

SOURCE

The following information has been collated from a consultation with the project leader, David L. Roberts, from the University of Kent.

TOOL DESCRIPTION:

The University of Kent have developed a tool to aid investigators and researchers looking at eBay. The aim of the tool is to use crowdsourcing to identify elephant ivory with participants being asked whether they think an item is made of elephant ivory and whether they can see the Schreger lines.

Using eBay's API, items are extracted based on specific search terms. The images are then presented to participants in the form of a 3x3 matrix of images in a similar format to a CAPTCHA test[21]. The participant then selects the images in the 3x3 that they think could be elephant ivory. Once this decision is saved each image is presented individually to the participant and they are asked if they can see Schreger lines in the image. Once complete these classifications are saved. Because of the potential challenges of identifying elephant ivory and Schreger lines a training video has been developed and throughout known images are injected into the system so that participants' responses can be weighted for accuracy.

While these classifications are being used to develop AI technology solutions, there is the potential to use this crowdsourcing system to provide classified items to law enforcement. For example, once an image has been classified six times by participants as elephant ivory and that the associated seller has ten classifications against them then this intelligence package could be passed onto law enforcement.

Finally, while the system has been developed for elephant ivory classification, as it is based on search terms, there is no formal reason why the system could not be duplicated for other wildlife products.

[21] Captcha is the acronym for Completely Automated Public Turing test to tell Computers and Humans Apart. The most common version distorts letters and numbers and asks users to interpret and reproduce these characters to prove that they are not robots. Several versions of this test exist, including based on image recognition.

CONCLUSION
AND RECOMMENDATIONS

CONCLUSIONS AND

RECOMMENDATIONS

This report sought to assess how technology, in particular that derived from work in the field of data science, could help in the fight against wildlife cybercrime. The potential of the solutions presented here to crack down on wildlife trafficking is promising. In recent years, computer science has moved forward rapidly and mastered algorithmic models that are able not only to identify wildlife sold on the internet but also to help assessing illegality. These solutions could possibly represent a substantial improvement in LEAs' capacity to quickly, and efficiently monitor actions related to the trafficking of protected species online.

However, at present there is no one-fit-for-all, reliable, systematic and repeatable way to detect wildlife cybercrime. Moreover, it is too soon to identify amongst the systems presented above the ones that really are going to be scalable, reliable, and cost-effective in a foreseeable future.

To fulfill this potential, the development of data science in the field of fighting wildlife cybercrime must go hand in hand with dedicated practitioners able to understand the phenomenon and monitor its evolutions in order to adapt data science programmes to wildlife cybercrime needs. Even though there is not yet any one-fit-for-all solution available, this report shows a growing availability of promising solutions, both in the academic and private sector, which should encourage the LEAs, the conservation organisations, and technology companies to keep working collaboratively and be creative in order to tackle wildlife cybercrime. The modus operandi of illegal traders evolves rapidly; as soon as understanding develops on some part of the phenomenon, it may evolve into something different. The only way to tackle this phenomenon is to put collaboration at the core of the development of effective tools.

Despite the appealing possible benefits of technology, combating wildlife cybercrime will require a multi-stakeholder collaboration and coordination. Effective collaboration, involving the relevant stakeholders, such as Police, Customs services, academia, NGOs and the giant tech-companies is required for automated solutions to be used to their highest potential.

Even though algorithmic models behind machine learning can be increasingly easily adapted and refined, the most important challenge lies no longer in the development of these technical tools, but in our ability to use their full potential, by feeding them the necessary data. CITES protects about 35,000 species, each can be characterized via different modes of communication both written (name of the species, paraphrase, slang name), or visual (photos, videos). The amount of data needed to enable computers to locate and identify these species is enormous, and difficult to compute. At the time of writing, no entity known to TRAFFIC, has gathered the amount of data required, and applied the technical means of analysing wildlife related data on a global scale. Only specialist technology companies, whose main activity is the acquisition and processing of data, are likely to have that potential. But the resources required to do this on a global level would be prohibitive based on conservation and enforcement budgets. However, targeted country or species-specific efforts are working, and with future advancements of technology and research, may be more widely applied in future.

Based on the research conducted in the context of this report, the LEAs are recommended to:

- ✓ **Build and centralise the critical datasets** necessary for scaling-up automated solutions;
- ✓ **Strengthen the collaboration between law enforcement agencies** on sharing best practices and knowledge on wildlife cybercrime;
- ✓ **Strengthen the internal collaboration between the cyber units and wildlife crime units** at the national level in order to tackle efficiently the multidisciplinary aspect of wildlife cybercrime;
- ✓ **Strengthen the collaboration with representatives of civil society** (NGOs and academia) to develop tailor-made tools;
- ✓ **Strengthen the collaboration with major players in the private sector** active in data collection and analysis to develop efficient tools and conduct decisive actions;
- ✓ **Strengthen internal skills and build capacity amongst the LEAs** to facilitate the integration of technological solutions into the daily activities of officers responsible for monitoring wildlife crime.

REFERENCES

Abbasi, A., 2012. Metafraud: a meta-learning framework for detecting financial fraud. MIS Quaterly, December, pp. 1293-1327.

Austen, G. E., Bindemann, M., Griffiths, R. A. & Roberts, D. L., 2018. *Species identification by conservation practitioners using online images: accuracy and agreement between experts*. PeerJ.

Butterfield, A., Ngondi, G. E. & Kerr, A., 2016. A Dictionary of Computing. 7th ed. *Oxford: Oxford University Press*.

Carl, M., Jack, P. & Josh, S., 2019. Detecting Online Environmental Crime Markets, s.l.: *The Global Initiative Against Transnational Organized Crime*.

Daly, N., 2017. Instagram Fights Animal Abuse With New Alert System. *National Geographic*.

Di Minin, E., Fink, C., Hiippala, T. & Tenkanen, H., 2018. A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation Biology*, 33(1).

Fink, C., Hausmann, A. & Di Minin, E., 2019. Online sentiment towards iconic species. *Biological Conservation*, 6 November.

Guan, J. & Xu, L., 2015. *Deadly Messaging: Ivory Trade in China's Social Media*, Cambridge: TRAFFIC.

Haysom, S., 2018. *Digitally enhanced responses*, s.l.: The Global Initiative Against Transnational Organized Crime.

Haysom, S., 2019. *In Search of cyber-enabled disruption*, s.l.: The Global Initiative Against Transnational Organized Crime.

Hinsley, A., 2018. *The role of online platforms in the illegal orchid trade for South East Asia*, s.l.: The Global Initiative Against Transnational Organized Crime.

IFAW, 2018. *Disrupt: Wildlife Cybercrime*, s.l.: s.n.

Lavorgna, A., 2014. Wildlife trafficking in the Internet age. *Crime Science*, 3(5).

Roberts, D. L. & Hernandez-Castro, J., 2015. Automatic detection of potentially illegal online sales of elephant ivory via data mining. PeerJ *Computer Science*, Issue July.

The World Bank, 2018. Tools and Resources to Combat Illegal Wildlife Trade, Washington: *The World Bank*.

Thomasz, F., 2018. Illicit wildlife markets and the dark web: A scenario of the changing dynamics, s.l.: *The Global Organization Against Transnational Organized Crime*.

TRAFFIC, 2019. *Combatting Wildlife Linked to the Internet: Global Trends and China's Experience*, Cambridge: s.n.

Wingard, J. & Pascual, M., 2018. *Catch me if you can: Legal challenges to illicit wildlife trafficking over the internet*, s.l.: Global Initiative Against Transnational Organized Crime.

Wingard, J. & Pascual, M., 2018. *Catch Me if You Can: Legal challenges to illicit wildlife trafficking over the internet*, s.l.: The Global Initiative Against Transnational Organized Crime.

Xiao, Y. & Wang, J., 2015. Moving Targets: Tracking Online Sales of Illegal Wildlife Products in China, Cambridge: TRAFFIC.

Xinhua, 2018. China ups fight against wildlife crime with tech partnership. *Global Times*.

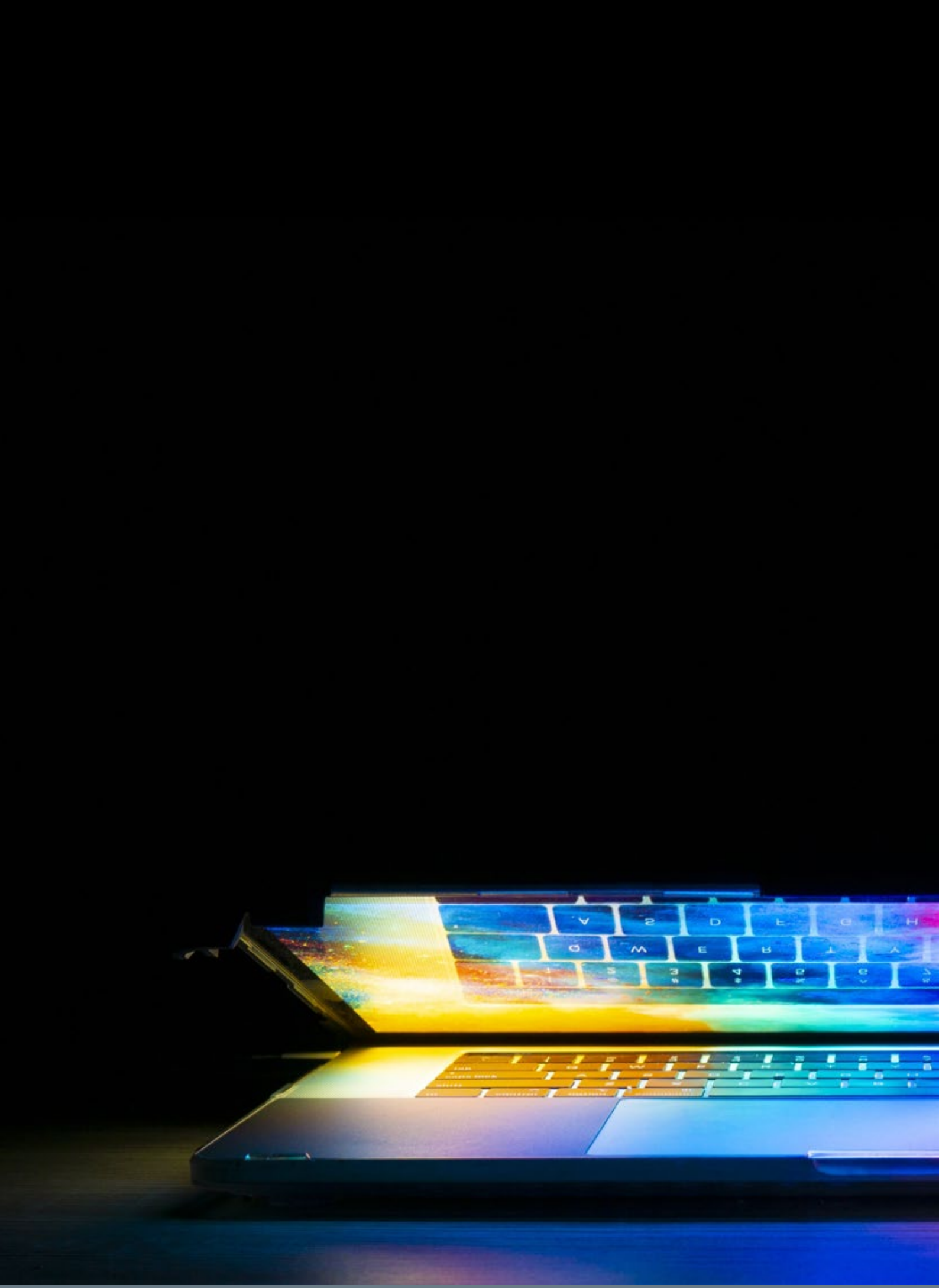
Xu, Q., Jiawei, L., Mingxiang, C. & Tim, M. K., 2019. Use of machine learning to Detect Wildlife Product Promotion and Sales on Twitter, s.l.: *Frontiers in Big Data*.

Yu, X., Jing, G. & Ling, X., 2017. *Wildlife Cybercrime in China: E-commerce and social media monitoring in 2016*, Cambridge: TRAFFIC.

IMAGE CREDITS

Unless otherwise stated all images are Creative Commoms 2.0 Non-attribution.

Page	Credit
16	National Geographic Stock



JULY 2020

This document has been compiled by WWF-Belgium with financial support from the European Commission and in-kind support from TRAFFIC.

WWF is one of the world's largest and most experienced independent conservation organizations, with over 5 million supporters and a global network active in more than 100 countries. WWF's mission is to stop the degradation of the planet's natural environment and to build a future in which humans live in harmony with nature, by conserving the world's biological diversity, ensuring that the use of renewable natural resources is sustainable, and promoting the reduction of pollution and wasteful consumption

TRAFFIC is a leading non-governmental organisation working globally on trade in wild animals and plants in the context of both biodiversity conservation and sustainable development.

For further information contact:
TRAFFIC
Global Office
David Attenborough Building
Pembroke Street
Cambridge CB2 3QZ
UK

+44 (0)1223 277427
traffic@traffic.org
traffic.org

UK Registered Charity No. 1076722,
Registered Limited Company No. 3785518.

European Commission

TRAFFIC
the wildlife trade monitoring network

WWF
Working to sustain the natural world for the benefit of people and wildlife.
together possible.