Darwin Initiative, Darwin Plus and Illegal Wildlife Trade Challenge Fund Covid-19 Rapid Response Round - Final Report

Due within two months of the end date of the Rapid Response Round project

(maximum 6 pages)

Project reference	CV19RR\1035
If linked with an ongoing project, please include that project reference here (e.g. IWT001)	
Project title	Uncovering the illegal online trade in South African succulents- 'Project Ogie'
Country/ies	South Africa
Lead organisation	Royal Botanic Gardens Kew
Partner institution(s)	South African partners: South African National Biodiversity Institute (SANBI), Department of Forestry, Fisheries and the Environment (DFFE), South African National Parks SANParks), CapeNature, University of Southampton (UoS)
Start/end date of project	1 Feb 2021 to 31 March 2021
Which fund was this project relevant to?	Darwin Initiative / Darwin Plus / Illegal Wildlife Trade Challenge Fund
Grant value (£)	£19,208
Project Leader name	Dr Carly Cowell
Report author(s) and date	Carly Cowell & David Whitehead

1. Project Summary

South Africa has 4337 succulent plants of which 566 are threatened and are listed on the Red List of South African Plants. Currently, approximately 810 species have trade listed as a threat to their survival in the wild. Poaching of South Africa's succulent plants and caudiform species has been escalating exponentially since March 2019 made worse by the COVID19 pandemic. As economies suffer from lockdowns, unemployment figures increase so the number of poaching incidents has increased as people look for alternative means of income (Curtis 2020). South African enforcement officials have seen that poaching has moved from international visitors doing the poaching themselves to these same buyers now soliciting from local South Africans. Illegal trade routes need to be identified and stopped and sustainable conservation friendly jobs created for those most affected.

Trade of South Africa succulents has rapidly increased as have poaching events and seizures of wild harvested plants species. Tens of thousands of plants are being confiscated by South

African law enforcers. The South African National Biodiversity Institute (SANBI) and the South African National Parks (SANParks) even suspect that some species in the genus *Conophytum* have likely been poached to extinction in January 2021 alone. Given that South Africa is a relatively remote country with vast areas of land to be policed for plant poaching, online Artificial Intelligence (AI) tools were employed to determine if they could help identify trade and demand.

As South Africa is documented as a current hotspot of plant poaching for the horticultural trade, focusing efforts on this region are likely to yield significant results, with potential implications for plant conservation both in South Africa, and within the destination countries where illegally collected plants may be trafficked to for sale.

The project combined innovative and cross-disciplinary ways of analysing online marketplaces to determine which species were in high demand and where potential markets were positioned (ie. destination countries).

The project focussed on the Succulent Karoo Biome region (Figure 1) of South Africa, an area of 102,700 $km^{2.}$ This biome occurs in the Northern Cape, Western Cape and Eastern Cape Provinces.

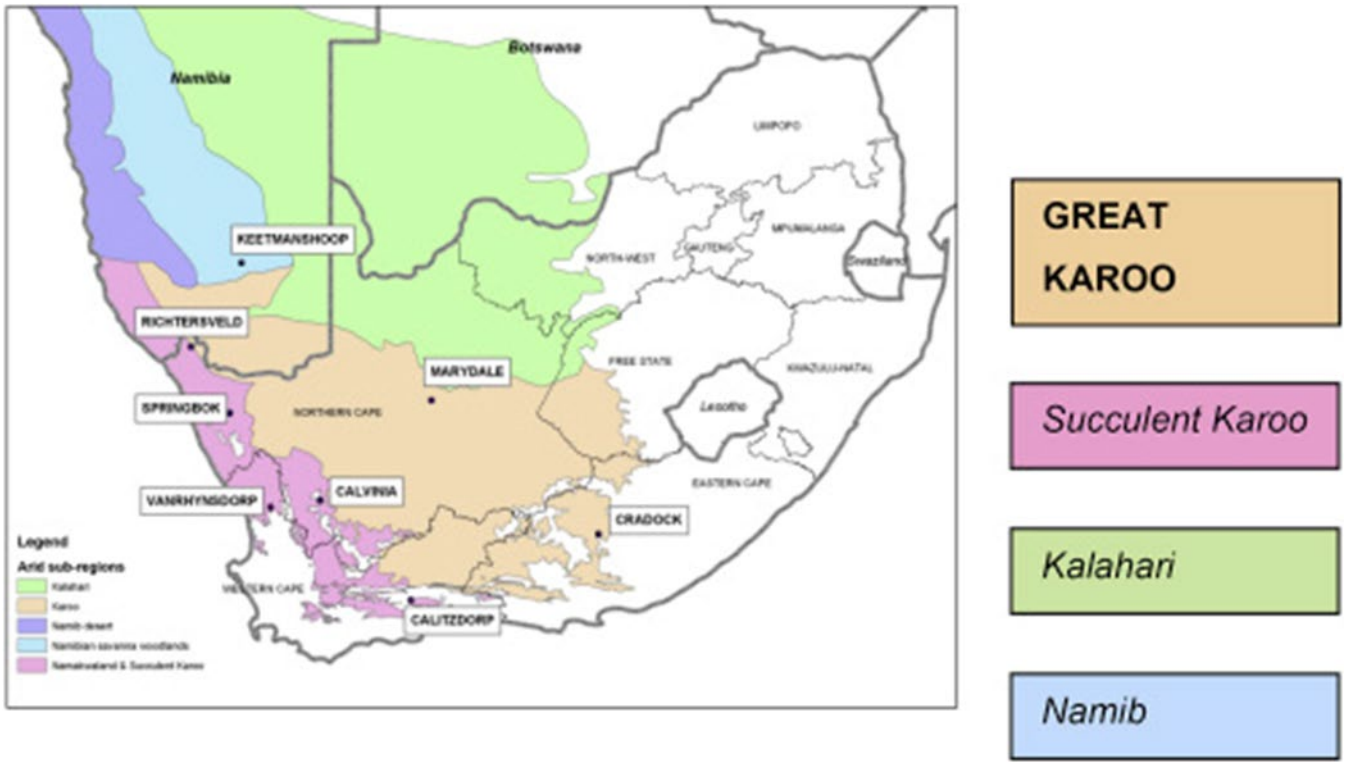

Figure 1. A map of South Africa, showing the nine provinces and highlighting the arid vegetation regions. Of particular interest to this project is the Succulent Karoo region in pink where much of the current succulent plant poaching is occurring.

2. Project Achievements

2.1 Intended Outcome

The project was designed to test the proof of concept that AI software, used by a cross-disciplinary team, could be used to automatically search pre-identified websites and forums for posts mentioning *Conophytum* spp and caudiciform species from South Africa. Posts matching a search lexicon of key terms would be extracted and analysed. The project further aimed to build capacity of the South African partner organisations by training staff to use the methodology and AI tools into the future to monitor trade and potentially prevent poaching events.

The four objectives of the project were:

A. Establish which species are frequently traded:

Target species were selected from species records of 2000 South African succulents supplied by SANBI. The first AI crawl was done using a search lexicon which included twelve of the top seized *Conophytum* species and caudiciform species supplied by the Department of Forestry, Fisheries and the Environment (DFFE). The second crawl looked at forums using the lexicon from crawl 1 and a list of forums identified by the project team, and the third crawl looked at marketplaces, using the same lexicon as in crawl 2.

Determine which countries South African succulents are being traded in:

The website domains were used to highlight which countries had ongoing trade (buying and selling) of South African succulents. Further analysis of returned data found that Asian and European countries were the most prolific in demand for succulent species from South Africa for sale.

B. Establish whether possible collection/poaching events are going to take place:

This objective was not achieved for the reasons stated above in Section 2.4, due to GDPR restrictions this was not possible in the given project time frame.

C. Build capacity in the UK and South Africa to use AI tools to monitor online trade in threatened species:

Three South African staff (two female) and one UK were trained on the use of the AI tools using a test forum created for training purposes. All software and programming information was supplied to partners and training on programming was given over a period of 4 intensive training sessions. Criminology analysis training for project partners was given to AI trained recipients and an additional three in-country partner staff, all of whom were from disadvantaged backgrounds. A general guideline document was produced to aid partners in future application of the methodology through a stepwise guidance process. A network of over 20 international and local experts was established to provide ongoing expertise and advice as reported in Section 2.2 below.

2.2 Main activities undertaken.

Weekly team meetings were held to liaise with project partners with regards to the training on the AI tools and implementation of the internet crawls. Monthly meetings were held with the larger project group of partners to discuss progress and updates on the situation on the ground in South Africa. Lexicons of key words for search input into algorithms were established, one for *Conophytums* and another for caudiciform plants. Two encrypted external hard drives were purchased for secure storage of raw data and potentially sensitive information. This ensured adherence to data protection regulations.

Training on the socio-technical Natural Language Processing (NLP) methodology was conducted online, in sessions spanning 4 days and around 12 hours total training. This was well received and the method was picked up quite quickly. Feedback after the sessions was good and provided enough information for iterative improvements to the AI tools to be made. The training provided important contextual information around the software, establishing good foundational knowledge across the group of trainees.

Contact was made with several international experts for advice on species and websites to use in the crawls. This allowed for a wider group of experts to be aware of the poaching problem and provide a louder ‘voice’ for anti-plant poaching in South Africa and support for communities via education and skills development. In order to develop a comprehensive search lexicon all partners made contact with known experts within their own networks and gathered information. This was invaluable as it provided expert local knowledge from those on the ground dealing with the effects of illegal plant poaching in South Africa. The international botanic garden network was also consulted for views on the species in trade, volumes and destinations of purchases, once again providing community insights to the botanical/horticultural trade network.

A secure space using MS Teams was setup with assistance from the IT department at RGB Kew for the project partners to share information, documents and where honest conversations and knowledge sharing can take place.

Individual agreements were not signed by experts as they were not needed, as project PI is a member of the IUCN Cacti and Succulents Specialist Working Group and therefore has access to experts within by this working group.

Criminology training took place on the 17 March 2021, for five South African staff and three UK staff. This will enable conservationists to identify suspect posts and behaviour more rapidly, and alert enforcement authorities much sooner, in future crawls in South Africa.

First online search and data collection for *Conophytum* spp and Caudiciform species.

Here the results from the crawl of only seven online sites (two forums and five market places) yielded a mix of results. The aim of this first crawl was to determine whether the lexicons produced were appropriate and whether the updated AI software could produce results similar to the FloraGuard Project. We found that the lexicons worked well and only the term 'window' when referring to a 'window plant' as a common reference to certain species produced false results and had to be removed from the lexicon. None of the marketplaces returned results on *Conophytums* for sale. The most popular South African caudiciform for sale are the succulent *Othonna* species (*Othonna glauca, O. euphobioides* and *O. herrei*). Other species which were frequently mentioned for sale were *Pelargonium triste* and *Tylecodon wallichii.*

Second online search and data collection: Forums relating to Conophytum and Caudiciform species

Seven English language forums of potential relevance to the target species were identified, either through manual searching for relevant threads, or based on their general topics of interest. After an evaluation of the terms and conditions, five were selected and successfully crawled, with high volumes of data returned. Use of a single search lexicon for all species improved the initial efficiency of the crawl, although produced a larger, combined data set for analysis. Five non-English language forums were also identified, one of which was crawled but did not return data.

Third online search and data collection: Marketplaces relating to Conophytum and Caudiciform species

One hundred potential websites were identified through manual inspections by the project team. These included English language (UK, USA) and non-English language marketplaces (Chinese, Japanese and Korea). In some cases, marketplaces originating from non-English speaking territories (Italy, Germany) were presented predominantly in English, without the need for a translation tool (which the crawler cannot automatically apply). These marketplaces ranged from specialist online nurseries to broad general sales platforms, containing subsections dedicated to plants. Following an evaluation of the Terms and Conditions, and closer manual inspection of the contents of each site, 15 were selected for crawling. Of these, six returned data from the crawls. In our approach, we used the combined lexicon of terms for the two plant groups used in the second crawl in order to compare results if possible.

A Handbook providing general guidelines on setting up the AI tools and the various aspects that must be considered when undertaking this type of work. This has been produced as a PDF document which can then be updated as needed and shared to a wider audience both in the UK and South Africa. The Handbook has been written in a sequential way and takes practitioners and users through a step-by-step process of how to get started and what is needed for a

successful application of this AI technology. Support is also on hand through the project partnerships and established network of contacts.

2.3 Highlights of work not originally planned

Due to delays in approval for the use of special information as regulated by UK GDPR, as a result of Covid19 furlough of essential staff, an online training forum was established. This was populated with fake posts and personal data (relating to imaginary “Unicorn trade”), to mimic the type of online discussion forum where rare plants are often mentioned, including in the context of sales. This forum was then used within the training sessions, to enable the full data crawl and analysis cycle to be performed. The training forum can now be used in the future to train staff on the use of the AI tools and analysis techniques without risk of data protection breaches or risk to staff. The need for a similar training marketplace site is now needed as we have found that forums and marketplaces tend to be designed differently and require special configuration of the AI tools respectively.

During the training, it was found that many sites selecting or discussing South African succulents are in languages other than English and this was identified as a gap in the existing AI tools. The socio-technical NLP methodology was updated so that it is now capable of crawling sites in different languages, although this still needs to be refined and native speakers will need to be available to assist to ensure accuracy. Further updates to the software included: the ability to download data into an excel format for ease of analysis and linking to the Named Entity Graphs; the addition of “date” as a specific piece of information recognised by the algorithm; and the inclusion of an “excluded terms” lexicon (to filter out false positive results), which were not features of the software package at the outset of the project. These updates will make it much easier for practitioners in South Africa to use the AI tools, and process and analyse large data sets.

2.4 Challenges

A Data Protection Impact Assessment (DPIA) was conducted at Kew, as the lead organisation for the project. Legal and personal data experts from Kew and external consultants were brought in to help with the interpretation of the GDPR as they apply to Kew and this project, as well as the UK and South Africa. Likewise, a legal advisor at DFFE, in South Africa was consulted regarding the POPI Act considering the use of AI tools to identify criminal activity related to illegal plant trade. The result was that trade and species data could be crawled by Kew, and criminal activities by DFFE, in future applications following training on the AI tools. The initial data crawled by Kew is currently being analysed to inform DFFE of species and countries to focus on in more depth.

Due to Covid19 and furlough schemes in both the UK and South Africa the project started a month late. Partners in South Africa faced problems with staff leaving or being transferred to other departments and those remaining having higher workloads. There was also the challenge of the increase in poaching events taking place in South Africa and many partners being involved in the sorting and identification of the plants seized. Here we scaled back the timeline for delivery of objectives to relieve the stress on project partners. Support was also given in the identification of species where possible.

Project partners were under immense pressure due to Covid19 workloads. Our computer science expert at the University of Southampton found that he could no longer assist with the training on the AI tools due to increased teaching requirements. This further delayed the start of the project and insufficient time was then had to run all 3 online searches. This was overcome by employing his top student who had worked on the FloraGuard project AI tools, as an intern at Kew, to do the training and provide support which was a game saver. Training of partners and online crawling was then conducted concurrently in order to maximise time. Online searches were continued in the extended project time and post-project period, into April and May 2021.

False positive results from the second and third crawls on Forums and Marketplaces was time consuming, but more of a challenge was errors produced form marketplace websites which appeared to be blocking the crawling although crawling permission were clear in the Terms and Conditions of these sites. Examining the difficulties encountered by the crawler with some of the marketplace websites, and how to best to engage with this type of website structure and the mix of languages presented, represents an important focus for future work. The establishment of a training marketplace site similar to the forum training one developed in this project is a possible solution.

2.5 How you contributed towards gender equality.

When considering the persons within the project team who would be trained on the use of the AI tools, it was considered a priority that women and people from previously disadvantaged backgrounds be given preference. This resulted in two female analysts and one male from a previously disadvantaged background in South Africa being selected. Further to this, the project team consisted of five women in senior roles who were provided with decision making skills and information to further their work in conservation. The women in senior CITES roles in South Africa and the UK felt more supported in their work on CITES by a broader international network.

The theory of ‘train the trainers’ was used in order for the skills shared in this project to be taken to a wider conservation staff component in South Africa. By providing the skills necessary to undertake the automated crawling of the internet and to provide further training to the South African staff, we were able to increase their profile within each of their workplaces. Feedback

from participants stated that they felt more confident in their roles as analysts and able to provide advice and support to others working on similar challenges with other species such as the trade of reptiles online.

2.6 How your project delivered value for money.

The entire project team did not have in-depth knowledge about personal data regulations relating to AI sourced data from the internet but as a result of the work on this project all have a much great understanding of the requirements and what is needed to take the next steps in order to use the AI tools to detect illegal activities. The skills provided through the training and the software will enable more efficient use of resources for monitoring of online trade in South African succulents which is a mammoth task currently requiring resources and capacity not available. It can also be expanded to other plant species and adapted for animal species or other taxa that are subject to illegal trade. The results of this project, and ability to conduct further analysis on the species in trade, will be used to inform Red List Assessments currently being conducted on South African succulents. This will reduce the research time to gather this type of data for Red Listing assessments and will aid South African CITES Authorities in the development of proposals for national consideration of species listings onto CITES. All of this work will continue to be fully supported by the UK project team, with any update of the AI software supplied for free to project partners. The project has also enabled the ability of conservation scientists to be trained in the software and operate it with minimal ICT support, to be assessed. The encouraging results will make wider deployment of the software in other settings more feasible and easier to implement going forward.

2.7 The effectiveness of partnerships.

A request to assist with data gathering of succulent species, particularly regarding trade data, was made by colleagues within SANBI to Kew in 2019. This data was to be used for Red List assessments of those species traded in high volumes or in high demand. Joint discussions followed the launch of the FloraGuard Project report and this project was then agreed on. Since writing and submitting the project proposal partnerships have strengthened and sharing of information and collaborations has improved over the period of the project. The websites targeted for crawling by the AI tool were guided directly by the South African partners particularly analysts from SANBI and DFFE. There are plans to continue this work after the end of this project and to pursue further funding for the training of more staff in other regions in South Africa, support for trained staff and further web crawls and analysis of data. Friendships and working relationships have blossomed and grown, as a result a much better network of collaborators and support is now available to conservationists in looking to stop the illegal trade in succulent species from South Africa.

All partners were involved in decision making throughout the life of the project, monthly meetings took place with the whole project team, weekly meetings with the research team and daily communication between the technical team. Select areas of expertise were allocated to various team members for their input or to gather data such as seizure data from CapeNature and DFFE enforcement teams and AI software and training by computer science experts at the University of Southampton. The final report was written by the project PI with input by partners where their expertise were needed. All team members had sight of the final report prior to submission.

2.8 Safeguarding and ethical considerations.

Due to the sensitive nature of the topic of this project and the potential for alerting illegal wildlife traders to the work, no information was put on social media or in the press. The security of project partners particularly in South Africa is a concern as there have been threats against those involved in apprehending and prosecuting plant poachers. For this reason, all the names and details of the project team have been kept out of the media. Also, important to note is that due to the nature of wildlife cybercrime, it was not possible to contact data subjects to provide them with a privacy notice (inform them of the searches taking place), this is known as informed consent. For the success of this type of work and achieving the aim of identifying possible poaching events in the future, one must consider if a request for informed consent would disrupt the natural flow of information on the site, with a possible increase of risk to researchers.

Although working relationships and partners in projects develop trust and work well together, there may be higher levels of agreement and trust that need to be considered prior to engaging in international projects. The sharing of possible personal data particularly with an aim of identifying potential criminal activity may not be possible between certain countries without international agreements in place. These take a very long time to be discussed and implemented at government levels and are often not practical for research projects seeking to curb illegal trade taking place. The guidance is therefore to include partners who already have such international data sharing agreements in place. The use of a training forum is another option whereby partners are trained in the methodology and can discuss, analyse and share data and there are no international risks or need for agreements at that stage.

Similarly, when conducting live online crawls of the web for illegal trade data, the data owner organisations (which will have the software and raw data downloaded to its computers) should have permissions in place allowing them to do so. Many national data protection laws (GPDR-UK and POPI ACT-South Africa) make provision for enforcement agencies to search for and act on suspect criminal intelligence data.

2.9 Contribution towards the wider aims of the Illegal Wildlife Trade Challenge Fund

There are plans to continue this work after the end of this project and to pursue further funding for the training of more staff in other regions in South Africa, support for trained staff and further web crawls and analysis of data. Expanding the use of the AI tool and methodology for other IWT species traded online is also now possible for project partners. Friendships and working relationships have blossomed and grown, as a result a much greater network of collaborators and support is now available to conservationists working to stop the illegal trade in succulent species from South Africa. As a result, the aim of strengthening law enforcement has and will continue to be achieved. Additionally, a deeper understanding of the trade in South African succulents will enable better engagement with local communities with regards to sustainable harvesting and conservation of the species in trade.

3.Lessons learnt

Cybercrime involving IWT is a serious and increasing problem globally. To understand the scale of this problem the right permissions are needed to search the internet for possible illegal activity. Data protection rights are in place to protect people, yet it is our moral responsibility as conservationists to ensure that plant species are not negatively impacted by online trade. Concurrently, it is also our moral obligation to ensure that our research does not infringe on privacy rights of online users. Ultimately, if online IWT cannot be effectively policed due to potential breaches of personal data rights, then efforts to combat IWT face a significant problem. This project made substantial progress in identifying where the constraints lie. There is now an opportunity for the IWT Challenge Fund programme to develop and support mechanisms that address these constraints to ensure a balance between privacy rights and the rights of species to avoid extinction.

Under the current situation of a global pandemic, it was a requirement that all project partners be flexible and adaptable to a constantly shifting situation with regards to Covid19 lockdowns, budget cuts and staff changes. Similarly, the administration of projects like this should be flexible and open to the shifting dynamics of situations beyond the control of a project.

Aspects of the project that worked well were, getting partners on-board. There was so much enthusiasm for the work that initially there were too many partners to include in such a project and we had to be strategic for security reasons and practicality. The setting up a training forum that would not breach any data protection acts but still allow for the training of partners to use the tools in a real, live setting, when national permissions are in place was a huge benefit. However, the time it took to obtain permissions to conduct research using personal data and the constraints of GDPR was a challenge. We found that older research institutions are at times adverse to change and embracing new ‘high risk’ technology. Behavioural change in risk adverse organisations takes time and in the context of this project and indeed IWT time is not something

that is readily available. We also found that data sharing between countries requires agreements and permissions to be signed at the highest levels and this is lengthy and bureaucratic at best, there is a need for international co-operation on cybercrime regarding plants and for data and technology sharing to be better facilitated.

Should this project continue, we would partner directly with a law enforcement agency in the UK as the lead, who could be the data controller for raw HTML crawled data containing personal data. The partners in South Africa have these permissions. There is also a possibly for NGOs to undertake this work as they could potentially be in a better placed position for international data sharing.

As a result of our learnings from this project, we recommend that similar projects partner with organisations that have clearance to conduct the research to the full scope of the AI tools. This will expedite the roll-out of the project. Bringing together organisations dealing with similar IWT issues, has strengthened national and international partnerships and built new ones for the future, as each one brings new and valuable information to the project. The use of a training forum was immensely valuable as it allowed for training to take place almost immediately whilst other permissions were still being obtained. The use of the training forum from this project is still available for others as the need arises.

The crawling technology remains in development, and the research performed for Project Ogie has helped to refine and test its application in a range of practical settings. Reflections and technical observations from the search of forum posts found that, configuring the crawler to forum sites proved relatively straightforward, with these sites tending to follow a more reliable structure, that fits well with the crawler's configuration parameters. There was one exception a non-English forum, where despite the occasional key word in English, the crawler was unable to return any results. The crawler does have functionality to search for single key words in other languages, and further work is required to develop a workflow incorporating translations of key words and results.

The capture of the creation date for each forum thread proved a useful feature of the crawler, and cases where this information was not returned can now be examined more closely, to further refine our website configuration techniques.

Crawling marketplaces proved considerably more difficult. Marketplace settings have a less reliable structure in terms of their navigation and menu settings, making it more difficult to successfully configure the crawler to capture all of the available data. Nine of the marketplaces crawled did not return any data, and examining and better understanding the structure of these tricky websites, will help greatly with the design of the next iteration of the software. The return of 404 errors from a key website of interest also requires investigation. Improving the

performance of the crawler in marketplace settings, particularly in non-English language website settings, probably represents the “next generation” of the software to strive for.

In a small number of cases, websites were deemed “uncrawlable” for other reasons. This category of “red flag” websites is characterised by amateur website designs, with the website typically containing a mix of discussion, links and sales content. This only affected three websites assessed during the course of Project Ogie and working on a version of the crawler to draw out key content from these oddly assembled sites is a consideration for further software development.

For better structured websites, the crawled results were markedly better. In some cases, the crawler was able to capture the price of each item within the data set, although in one case these were uniform values that did not reflect the website content, although this may be more a function of the website, than the operation of the crawler itself. The addition of false positives to the data set only affected a few of the crawled websites, and did not significantly hinder data analysis. Carefully cross-referencing these posts with our search lexicon (which contained a total of 70 terms) will determine whether these are genuine false positives, or whether the crawler was capturing them in relation to the lexicon in a less obvious way.

Last, but not least, was the opportunity to prove whether, following initial training, a conservation scientist was able “fly solo” with the crawling software, with minimal support from ICT experts. Here, the results were pleasing, with website configurations set by both the software developer, and conservation scientist, used to successfully crawl the websites of interest. Additional experience in deploying the software, and reflecting on our results with ICT project partners, will no doubt improve crawling success further, but the concept of creating a tool that can be put to practical use by non-ICT specialists, has shown considerable promise.

Finally, one of the key lessons taken away from this project are the need to place GDPR requirements in institutional and country settings and have any issues addressed prior to commencement of the work. Older research institutions and organisations may not be familiar with AI tools and technologies and time needs to be spent working with relevant staff to explain processes and procedures to ensure legality of all work undertaken. Once this is done it will provide firm foundations to build on going forward for future projects of this nature.