

RESEARCH PAPER

Automatic Acquisition and Sustainable Use of Political-Ecological Data

Timothy C. Haas

Lubar School of Business, University of Wisconsin at Milwaukee, Milwaukee, WI 53201, US
haas@uwm.edu

The sustainable management of anthropogenically-impacted ecosystems will require ongoing monitoring and advocacy by people across the globe. To this end, automatic methods are developed herein for acquiring several types of such political-ecological data. On the political side, a method is developed for gathering news articles about human actions that affect the ecosystem along with a method for identifying themes in social media that concern the consumption of an ecosystem's products. On the ecosystem side, a method is derived for estimating wildlife abundance from purchasable high-resolution satellite images. A simple website architecture is described for holding this data and enabling its use in developing sustainable conservation policies. A rhino conservation website illustrates this architecture. A fundamental contradiction between the desire for open data on the locations of endangered flora and fauna versus the need to hide these locations from poachers is addressed through a new security protocol that enables the secure distribution of sensitive ecosystem data to trusted data consumers.

Keywords: socio-ecological data; environmental management; information extraction; mining social media; remote sensing; rhino conservation

1 Introduction

Humans are probably causing the earth's sixth mass extinction (Ceballos et al. 2015). Political actions play a large part in setting in motion the anthropogenic forces that are driving this destruction. But what is the nature of a process formed by political actions driving and being driven by ecosystem state variables such as animal abundance? Can such a *political-ecological* process or system be managed so that the affected ecosystem is sustainable? Any research agenda designed to address these questions needs a stream of jointly-observed political actions and ecosystem states. A political-ecological system is also referred to as a *socio-ecological* system or *social-ecological* system (e.g. see Virapongse et al. (2016)). The term, "political-ecological" is used here because political actions are often the precursors to social movements, and many social actions are initiated by groups seeking to increase their political power (Guinote 2017).

An associated issue is the need to make access to political-ecological data easier. For instance, the U.S. National Science Foundation (2014) states that

Researchers, policy-makers and others need access to well-described and easily discovered Earth observational data. These data form the basis for informed decision-making and wise management of resources. As Ecology evolves into a more data-intensive science, the ability to discover, integrate and analyze massive amounts of disparate information becomes critical, alongside a requirement to equip researchers with the skills to manage data effectively.

But data on the decisions humans make to manage ecosystems also needs to be easily accessible in order to assess the effects of management policies on those ecological processes being managed. Within the particular sustainability challenge of preserving an ecosystem's biodiversity, Rissman & Gillon (2017) believe that

Clear links between ecological and social dynamics, feedbacks, and outcomes are needed to effectively inform policy and management and improve the social-ecological fit of conservation strategies (Bodin et al. 2014).

Laurila-Pant et al. (2015) develop a framework for incorporating biodiversity protection into ecosystem management policies. But they warn

A thorough analysis using the suggested framework requires considerable multi-disciplinary data or modelling results from both the ecological responses and the economic value of biodiversity as well as the costs of implementing the management measures.

This literature then, points to a fundamental role that political-ecological data will play in efforts to build decision support tools that are specific enough and reliable enough for policymakers to implement and defend policies that aim to protect biodiversity. Therefore, the importance of this data-need is difficult to overstate since acquisition and access to political-ecological actions is a necessary precursor for developing theories of political-ecological systems, and decision support tools for managing them.

There is also a need for more involvement across all stakeholders in the management of at-risk ecosystems. Such a need is the motivation behind calls for more *citizen science* (see Newman et al. (2017)). This involves enabling individuals to acquire ecosystem data themselves rather than these citizens attempting to access data collected by members of scientific elites such as those connected to large, funded research projects.

To address such data-needs, this article describes methods for automatically acquiring data on political-ecological systems, and a series of protocols for this data's safe-keeping and secure distribution. The data's structure is described along with specialized software needed to process it, and needed physical infrastructure to support its dissemination and use via the World Wide Web. The specific contributions that this article makes to the practice of sustainable ecosystem management are:

1. an original taxonomy for organizing political-ecological data,
2. a new product themes discovery algorithm,
3. a new megafauna abundance estimator,
4. a new data release security protocol, and
5. a website architecture that organizes political-ecological data in a simple and immediately useful way.

In the course of this description, a fundamental challenge in data science is addressed: how to provide access to data on an at-risk ecosystem for research and sustainable management purposes while at the same time keeping such data from those who would use it to harm the managed ecosystem.

Such data consists of both the actions of humans that directly and indirectly affect the managed ecosystem, and observations on metrics used to monitor the ecosystem's status. Examples of the latter include flora and fauna abundance and distribution; the distribution of toxic chemicals in the ecosystem's soil, air, and water; and the frequency and location of poaching events. Such a data repository will need to be updated in near real-time and needs to be available to those intent on ecosystem sustainability. Currently, environmental/wildlife protection agencies or projects funded by research grants provide some aspects of what an ideal data repository would have.

These current approaches to gathering and processing political-ecological data are viewed here as insufficient for two reasons. The first is that when struggling against political, budgetary, and corruption challenges, environmental protection agencies, no matter how dedicated individual staffers may be, can act as bottlenecks to the flow of political-ecological data into the hands of ecosystem management analysts and decision makers both within and outside such agencies. Instances of data withholding include South Africa's reluctance to release rhino poaching statistics (Legalbrief 2017), and China's editing of environmental data (Brombal 2017). The second reason is the short term duration of research grants. Rarely do research grants that have an ecosystem data collection component span more than five years. On-going political-ecological data collection, however, is crucial to learning about how ecosystems respond to anthropogenic inputs, and how such knowledge can be used to advance sustainable management policies.

The end result of such bottlenecks and short environmental study project schedules is few publicly available, long-duration political-ecological data sets and related analyses of candidate ecosystem management policies relative to the number of at-risk ecosystems. Because ecosystems function across international boundaries for time periods longer than those of political regimes, political-ecological data acquisition

and dissemination systems need to be able to operate regardless of whether ecosystem-hosting countries support or approve of the data collection and display program – and need to be inexpensive enough and distributed enough to continue functioning when funding agencies withdraw their support.

Such a publicly accessible political-ecological data resource is one component of an *ecosystem management tool* first proposed by Haas (2001, 2011: 59–78, 161–177; 2018). This tool is envisioned as a *global public good* as defined by Uitto (2016) in the context of environmental monitoring programs. In this view, knowledge of the state of the planet is a free good such as for example, language. Language is a public good (Gazzola 2014, ch. 3) and is crucial for human survival (Hoff 2003: 177; Morgan et al. 2015). Language is not owned by anyone but is used by everyone. If language was owned by some group and only those paying a fee were allowed to use it, civilization would quickly grind to a halt. Likewise, in order to sustain the planet, its managers (all humankind) need to know in real-time how their actions are affecting – and will affect critical ecosystems. Such knowledge then, needs to be accessible to all of the planet's stakeholders. Such real-time knowledge made available to all through an ecosystem management tool would be materially different from the current situation: finite environmental resources being consumed without any over-arching management policy wherein local management decisions are made to satisfy both local and geographically remote consumers – neither of whom have detailed and current knowledge of the effect that their consumption is having on ecosystems.

There is a computing aspect to data science (Blei & Smyth 2017). Computing may be as straightforward as standard summary statistics of a data set or as computationally demanding as fitting a model to data with Markov Chain Monte Carlo. Hence, a data portal realizes only part of the data science promise: the acquisition and use of data to expand knowledge. For the particular case of ecosystem management, this computing aspect of data science involves three main activities: (1) automatic acquisition and standardization of political-ecological data, (2) the fitting of statistical models to such data, and (3) computation on these fitted models to develop ecosystem management policies. Several techniques for the third activity exist such as *adaptive management* (see Rist et al. (2016)), and computation of the *most practical ecosystem management plan* of Haas (2011: 43–58). Because each of these three activities can be computationally demanding, developing an ecosystem management policy may require access to high-performance workstations or cluster computers. This article, however, will focus on algorithms for the automatic acquisition of political-ecological data, and the distribution of such data to trusted data consumers.

Reviewing the idea of sustainable development, Haas (2001) notes that long-term protection of an ecosystem is furthered with management policies that are (a) effective at protecting the ecosystem, (b) supported by all ecosystem-affecting countries, and (c) as economically efficient as possible. Organizations having the authority and resources to implement policy, could use a tool that can identify such sustainable policies. But who are these organizations? Currently, most environmental policy is developed and enforced by environmental protection agencies and wildlife protection agencies. Given the magnitude of their responsibility relative to the often limited financial support they receive, any management aid that does not come with a high price tag would be potentially useful to them. But there is another, often overlooked group of potential users: for-profit firms. Because for-profit or subsistence activities affect so many of the Earth's ecosystems, the involvement of for-profit firms is crucial to long-term environmental protection. For example, Lenzen et al. (2012) find that international trade and specifically the thousands of supply chains that firms in developed countries use to build and sell their products constitute the biggest threat to biodiversity.

In a potentially major advance in the management of the planet, for-profit firms would use an ecosystem management tool along with a separate computing resource to conduct their own assessments of different business actions open to them such as marketing an eco-friendly product, modifying their supply chain, underwriting a protected area, or, most impactfully, opening a service/manufacturing facility in an area where the local population is struggling economically. This last option would give local people economic options to ecosystem-damaging activities such as land clearing or the poaching of rare plants, hardwood trees, birds, amphibians, terrestrial animals, or aquatic animals.

In sum then, an ecosystem management tool enjoys (a) permission-free operation, (b) inexpensive operation, and (c) programmable operation (so that the system may be automated). All data and code that constitutes the ecosystem management tool is mirrored at several worldwide locations so that the system is protected against hacking or diplomatic demands that it be shutdown due to its data or conclusions reflecting negatively on one or more groups or countries. In particular, because many endangered species have home range in developing countries, ecosystem management tools are needed that focus on ecosystems within such countries.

To illustrate a few of the types of political-ecological data that should be part of an ecosystem management tool, three such data types are delineated herein along with associated data collection software. These data types are political action observations scraped from the World Wide Web, themes surrounding wildlife products derived from social media posts, and observations on animal abundance derived from remotely sensed images. In addition, an ecosystem management tool focused on rhino conservation in South Africa is given as an example of one way to construct an ecosystem management tool for an ecosystem that is inside one or more developing countries.

1.1 Automatic acquisition of political actions data

Acquiring observations on human actions from news outlets is a form of *information extraction* (see Piskorski & Yangarber (2013)). An automatic system is given for acquiring ecosystem-affecting human actions from online news stories. This system combines web-crawling with a natural language parsing algorithm to extract reports of human actions that pertain to an ecosystem that is to be managed. The parsing step of this system is accomplished with a modified version of the memory-based, shallow parsing algorithm of Daelemans, Buchholz & Veenstra (1999).

Most parsing algorithms recognize several types of verbs. Single-word verbs can be either *regular* or *irregular*. And a *multi-word* verb uses more than one word to convey its meaning, e.g. "picked up." See British Council (2017) for a review of these types. Here, all of these types are subsumed under the term *m-word verb* where *m* is a positive integer. The Daelemans, Buchholz & Veenstra (1999) algorithm is modified by replacing their word part-of-speech lexicon with a stored lookup table of action-related *m-word* verbs, direct-object phrases, and prepositional phrases. See Aarts (2011) for definitions of these phrase types.

Human actions that can affect an ecosystem range from direct, physical contact with the ecosystem such as poaching plants and animals – to more indirect actions such as setting aside land for a wildlife reserve. Because human actions are embedded in political processes (such as laws defining what poaching is, what constitutes ecosystem protection, and what landuses are allowed) hereafter, human actions that affect an ecosystem will be referred to as *political actions* to emphasize this political milieu that ultimately characterizes all ecosystem-affecting human actions.

A more free-form approach to information extraction is to mine text for topics for which people have sentiments toward. People now often express their sentiments on social media such as Facebook™ or Twitter™. Algorithms have been developed to discover these topics and associated sentiments. One such algorithm is given herein and, as an example, used to extract topics and sentiments surrounding ginseng root (Panax) consumption.

1.2 Automatic acquisition of ecosystem data

Remote sensing of ecosystem variables can be an inexpensive way to regularly acquire some information on an ecosystem's status. Many countries allow their citizens to purchase satellite image products from private firms, hereafter referred to as *image providers*. Small spatio-temporal patches of such products have prices similar to those of personal computers. Using such images to extract data on ecosystem metrics is one way to break the data bottleneck referred to above. An algorithm is developed herein to estimate abundance of a pre-determined species of megafauna using two satellite images taken sequentially in time. An example of its use is shown.

A fundamental contradiction, however, arises when proposing to publish the locations plants or animals belonging to endangered species. On the one hand, the central goal of a public resource of political-ecological data is openness. On the other hand, making these locations public can lead to the loss (through poaching) of the very plants or animals the site has been set up to protect. In other words, there is a challenge to make such sensitive data accessible to those who intend to use it to further the ecosystem's sustainability but keep it from those who, intentionally or not, would use it to degrade the ecosystem's sustainability.

Based on these considerations then, an ecosystem management tool needs two types of security protocols to protect its data assets: one to safeguard the data against theft, called here a *data security protocol*, and one to protect against the dissemination of the locations of plants or animals belonging to endangered species to poachers. This latter protocol is referred to here as a *data release security protocol*.

The distinction between these two intended uses of ecosystem data is sometimes difficult. For instance, based on modeling, a data requestor may have concluded that culling a wildlife population in a certain ecosystem is the best way to make that population sustainable. Should this requestor be allowed to access data on the locations of plants or animals belonging to endangered species to aid their culling operation? A new data release security protocol is developed herein that uses a combination of expert judgement, a video call,

and advances in cryptography to negotiate these sometimes complicated determinations in order to limit the release of sensitive ecosystem data to only those intending to promote the ecosystem's sustainability.

1.3 Previous work

Several systems have been created for automatic acquisition of political events (see Schrodtt & Van Brackle (2013)). These systems use software to parse news stories, and then associate each story with a member of a political action ontology. These systems have been developed for events surrounding international relations and terrorism. To this author's knowledge, acquisition of joint observations of political events and associated ecosystem responses is not present in the literature.

Michener (2015) lists 19 data portals to ecological data. One of these is the well-known Long Term Ecological Research (LTER) portal (LTER 2017). LTER consists of data from mostly fixed sites in the U.S.. LTER is not intended to focus on a specific ecosystem let alone a political-ecological system. Rather, LTER is a depository for data from individual, unconnected research efforts that are funded to answer specific ecological questions rather than to build a picture of a single ecosystem through time. This is in-part a result of the short-term nature of most ecological research grants as mentioned above. LTER avoids the risk of having the locations of endangered species members fall into the hands of poachers by simply not placing such data on its portal, see Cedar Creek (2018).

In general, current data portals have dealt with the issue of releasing sensitive data by simply not doing so. As another example, the access page for survey data on endangered plant species in the Gammon Ranges National Park in Australia provided by the TERN data portal contains the following note:

```
Project Completeness
Restricted - Species data: Observational data of sensitive species
is obfuscated.
(Gammon Ranges 2017).
```

In summary, current ecological data portals are mainly repositories for individual research projects as opposed to projects that follow a single ecosystem over a long time interval. And, in lieu of applying a data release security protocol that restricts release of sensitive data for sustainable uses only, these portals avoid the risk of releasing such data by simply denying access to it.

A static, one-time collection of social-ecological data on the Brazilian Amazon is described in Lima et al. (2016). To this author's knowledge, however, other than the work reported on herein, there is no literature and no exemplars of websites that continuously gather political-ecological data, offer it to the public through security protocols, and offer software for using its data to compute ecosystem management policies.

1.4 Article layout

The remainder of this article proceeds as follows. In Section 2, as a specific example of automatic acquisition of political data, a new method is described for acquiring stories about political actions that affect an ecosystem along with a new method for discovering themes contained in social media posts about wildlife products. For illustration, political actions surrounding the poaching of the white rhinoceros (*Ceratotherium simum*) in South Africa are automatically acquired, and the product themes discovery algorithm is used to detect themes surrounding the use of the ginseng root – a plant currently experiencing poaching pressure. Then, as a specific example of automatic acquisition of ecosystem data, a new satellite image-based method is described in Section 3 that can inexpensively acquire spatio-temporal data on those ecosystem metrics that can be remotely observed. This algorithm is used to estimate the abundance of elephants inside San Diego Zoo Safari Park in Southern California, United States. In order to show how these three types of data could be collected and distributed, the workings of an ecosystem management tool's website is described in Section 4 along with an example focused on the conservation of white rhinos. The white rhino is facing poaching pressure driven by the illicit trade in rhino horn. I discuss issues surrounding political-ecological data acquisition and associated data release security protocols in Section 5. I reach several conclusions concerning the future of such systems in Section 6.

2 Acquisition of Political Actions and Product Themes

2.1 Acquiring political actions stories from online sources

An ecosystem management tool needs a constant stream of data on political actions. The frequency of such actions affecting any given ecosystem is usually too great for human coding to be used. This is because humans take too long and are too expensive to have them continuously reading and coding online news

articles about a particular ecosystem. Indeed, an entire field of machine coding of political events has developed to address this speed and cost bottleneck in the acquisition of political event data. See Schrodt & Van Brackle (2013) for an overview of these efforts. Therefore, here, an automatic method is developed for acquiring political actions that relate to a particular managed ecosystem. This method is a new *text mining* (see Ignatow & Mihalcea (2018)) method that is based on a *taxonomy* of political-ecological actions that is described next.

2.1.1 Political-ecological actions taxonomy

Frey and Cox (2015) call for ontologies to be developed so that a common language of social-ecological systems can emerge. In line with this call, an original taxonomy has been developed for ecosystem management actions called the Ecosystem Management Actions Taxonomy (EMAT) (see Haas (2011: 123–141)). An ontology is a special case of a taxonomy (American Society for Indexing 2018).

Each action in this taxonomy has been parsed into three *equivalency phrase sets*: *m*-word verbs; direct object phrases; and prepositional phrases. **Table 1** gives these sets for the EMAT action *arrest_rhino_poachers*. The EMAT consists of the following action/message categories: military, diplomatic, economic, environment, and ecosystem. The first three of these form the Behavioral Correlates of War (BCOW) taxonomy of Leng (1999). See Leng & Singer (1988) for the taxonomy's design goals, and Schrodt & Van Brackle (2013) for the position of BCOW within the literature of automatic acquisition of political events data. The fourth category consists of actions taken by humans that are directed at the environment, e.g. poaching. The fifth category consists of actions taken by non-anthropogenic actors such as elephants trampling crops.

To improve the parsing algorithm's efficiency, if an EMAT action's *m*-verb word equivalency phrase set contains a 1-word verb that is regular, then all conjugated forms of that verb are also included in the set. For example, the verb "arrest" is listed along with the words "arrests," "arresting," and "arrested."

2.1.2 Phrase similarity measure

A subsequence of *n* words in a natural language phrase is commonly referred to as an *n-gram*. One way to define the degree of similarity between two phrases is with a modified version of the *Word Simple N-gram*

Table 1: Phrase equivalency sets for the EMAT action *arrest_rhino_poachers*. Phrases are arbitrarily ordered within columns.

<i>m</i> -word verb	direct object phrase	prepositional phrase
arrest	the duo	with the horns
arrests	a series	in possession
arresting	charges	of charges
arrested	suspects	in courts
caught	cases	of rhino poaching
faces	discovery	of rhino
facing	rhino poacher	of poaching rhino
faced	rhino poachers	of illegally hunting a rhino
apprehends	poachers	of killing a white rhino
apprehending	alleged poachers	of freshly killed rhino
apprehended	surviving poachers	of some rhino horns
nab	suspects	of two rhino horns
nabs	men	for poaching rhino
nabbing	on suspicion	by the team
nabbed	in possession	
appear	after being found	
appears	including attempted poaching	
appearing		
appeared		

Overlap measure (see Cordeiro, Dias & Brazdil (2007)). Letting N be the length of the shortest phrase between phrases ph_1 and ph_2 , a modified version of this measure is:

$$SIM(ph_1, ph_2) \equiv \frac{1}{N} \sum_{n=1}^N \frac{C_{match}(n\text{-gram})}{C_s(n\text{-gram})} \quad (1)$$

where $C_{match}(n\text{-gram})$ sums the Levenshtein Distance (Levenshtein 1966; Yujian & Bo 2007) between all possible pairs of n -grams where a pair consists of one n -gram from ph_1 and the other from ph_2 . The quantity $C_s(n\text{-gram})$ counts the number of n -grams in the shorter phrase. Note that this measure lies in the unit interval.

This phrase similarity measure is used to compute the similarity between an EMAT action equivalency phrase and an observed phrase from a news story (hereafter called a *trial phrase*) as long as at least one of these phrases consists of two or more words. Otherwise, Levenshtein Distance is used to compute similarity.

2.1.3 Web scraping

Web scraping is the process of automatically acquiring information from the World Wide Web. See Vasani (2014) for a review. Due to tightened security, most news aggregators do not allow scripts to scrape stories from their sites unless the script owner is a subscriber. Subscriptions are typically not free. One inexpensive aggregator is Microsoft's Bing™. This article's ecosystem management tool website (see Haas (2018) contains two approaches to web scraping: one coded in VBscript™, and one coded in JavaScript.

Either approach to web scraping uses the same method for acquiring the raw HTML of an online news story about the ecosystem being managed. This method involves scheduling a script to run once a week at a specific day and time. This script executes the following four steps. (a) launch a web browser and load the news aggregator's webpage; (b) load a search string such as "Kruger rhino poaching" into the search dialog box and issue a search request; (c) load the webpage of each link (story) returned by the search; and (d) after a story loads into the browser, append its HTML contents to a local file along with a story identifier number.

2.1.4 Actions extraction algorithm

The algorithm for extracting EMAT actions from each story in this HTML file of scraped stories is as follows:

1. Extract the text portions of a story with the text extraction method that is part of the `jsoup` text processing system (Hedley 2017). This method only retrieves phrases that can be parsed following the rules of English grammar. Hence, text content of a webpage is effectively defined as those sequences of tokens that can be parsed under this grammar.
2. Search each sentence for up to two EMAT actions. Within a sentence, check for each EMAT action as follows. First, search for one of the action's equivalent m -word verbs. If found, search independently for a prepositional phrase and the m -word verb's direct object phrase across that action's associated equivalency phrase sets. Declare a match between a trial phrase and an equivalency phrase if $SIM(ph_1, ph_2) > 0.8$. Then, declare a match to this EMAT action if either a direct object phrase is matched, or a prepositional phrase is matched and the direct object phrase similarity value is above 0.5.

The algorithm's EMAT action search strategy allows phrases to appear in any order within a sentence. The three phrase searches are run simultaneously on different threads within the JAVA™ program that implements this algorithm (see Haas (2018)).

2.1.5 Learning

New equivalency phrases of an existing EMAT action are learned semi-automatically with a two-step procedure as follows:

1. A story is printed out for which an m -word verb of an EMAT action was found but for which direct object and/or prepositional phrase similarities were too weak to declare an EMAT action match.
2. This story is read by a human and if an EMAT action appears to be described in a new way, the story's direct object phrase is added to the associated EMAT action's set of direct object phrases along with (possibly) the story's prepositional phrase.

The algorithm's step involving a human makes the algorithm semi-automatic and dictates the algorithm will run no faster than the time needed by the human to read the story and identify any new equivalence set members.

Because keywords have been used to identify online news stories for downloading, there is reason to believe that each downloaded story has at least one ecosystem-relevant action in it. Hence, an algorithm is needed that attempts to discover a potentially new EMAT action in each story that failed to produce a match to the current collection of EMAT actions. One such algorithm is as follows:

1. Search for an EMAT action. If no action is found, print the first four non-common single-word verbs in the text that match verbs contained in a lexicon of non-common verbs. In other words, when a story is encountered for which no action can be found, sentences from the story that contain verbs from a pre-determined list of relevant verbs are printed out. A list of 150 irregular verbs taken from Reverso (2017), and a list of 600 regular verbs taken from EnglishClub (2017) are used to conduct these new-verb searches.
2. Use this printed information to decide whether the story's action should become a new EMAT action.

2.1.6 Application to rhino conservation

The political actions extraction algorithm is applied to the political-ecological system of rhino conservation in South Africa. Online news stories were searched over the period January 2011 through June 2017. This data is summarized in **Table 2** and plotted as a time series in **Figure 1**.

2.1.7 Performance evaluation

Two criteria that a political actions extraction algorithm should be evaluated on are its accuracy and speed. The algorithm's accuracy can be assessed by comparing the actions extracted from a random sample of stories to those extracted by a human reading the same set of stories. Using the set of human-extracted actions as the benchmark, the algorithm can make two types of errors. The first error is failing to extract an action in a story, and the second is extracting an action that does not exist in the story, referred to here as an *artificial action*.

As a preliminary evaluation of the algorithm's accuracy, a random sample of 18 stories was drawn from the 205 rhino conservation stories. A human found 24 actions in these 18 stories while the algorithm correctly found 21 actions, missed 3 actions, and found 3 artificial actions. The fraction of actions correctly extracted is 87.5%, and the ratio of artificial actions to human-extracted actions is 0.125. This last statistic can be interpreted as the number of artificial actions per human-detected action. This preliminary evaluation suggests the political actions extraction algorithm is fairly accurate at extracting political actions from the web.

On average, the algorithm processes a story in about 12 seconds on a 3.2 GHz Intel personal computer.

2.2 Detecting themes around wildlife products

Another type of information that can be extracted from online sources is data on sentiments expressed towards aspects of the ecosystem such as its wildlife, wildlife products, poaching, and anti-poaching measures. For example, posts from Facebook and Twitter made by Asians in the market that mention the topic of rhino horn could be used to discover what other topics and/or sentiments are associated with rhino horn. In addition, because rhino poachers often maintain a social media presence, sentiments towards the perceived risks and benefits from rhino poaching could also be collected from individuals living in areas that harbor poachers.

Detecting sentiments that consumers hold towards consumption of a wildlife product in a way that is not affected by well-known survey biases (see Nuno & ST. John (2015)) could prove to be an accurate way to determine the effectiveness of wildlife product demand reduction campaigns (see for example,

Table 2: Summary measures of the political actions data set acquired for the rhino conservation political-ecological system.

Summary measure	Value
Number of stories	205
Number of unique EMAT action types	4
Number of rural resident action types	1
Number of rural resident actions	111
Number of anti-poaching unit action types	3
Number of anti-poaching unit action types	94

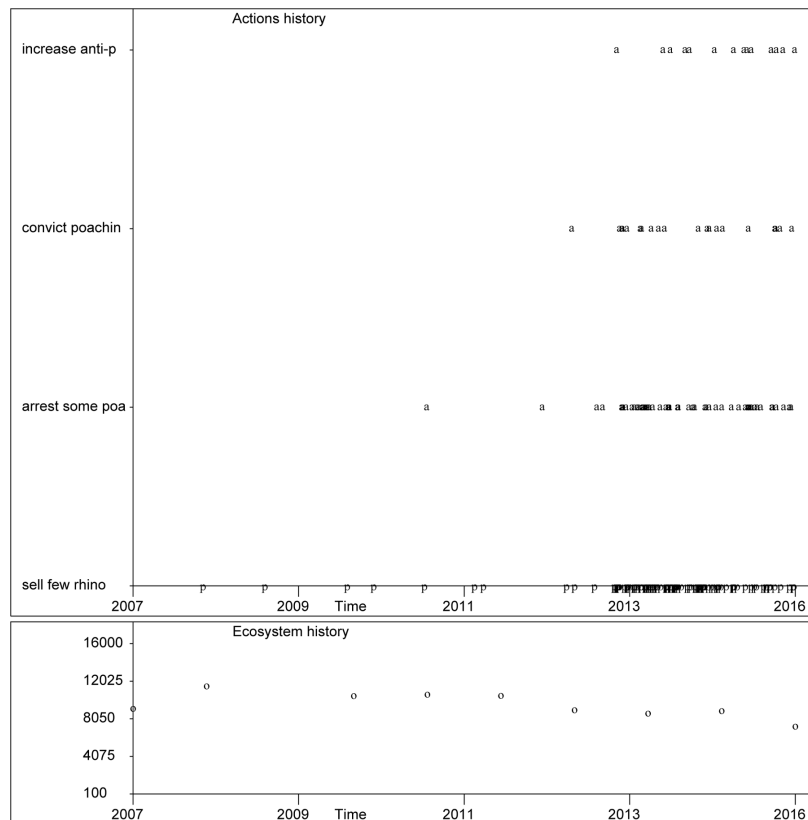


Figure 1: Political actions observed between 2007 and 2016 that are related to the rhino conservation political-ecological system. A plot symbol is the first letter of the group executing the output action: (p) oachers, and (a)nti-poaching forces. Estimates of rhino abundance in Kruger National Park, South Africa appear in the lower plot.

United States (2016)). It is possible, however, that consumers do not have a positive opinion of a wild-life product but rather view it as a necessity. But typically, sentiment analysis implicitly assumes that social media participants have a positive-to-negative valence towards a product. See for example, Lim & Buntine (2014). For a relatively unexplored product such as rhino horn, it may be more useful to first discover what characteristics consumers attach to the product through an exploratory analysis of social media postings concerning it. If such a positive-to-negative valence does appear, then a more accurate Method to ascertain the strength of that sentiment may be applied such as the algorithm of Lim & Buntine (2014). Otherwise, demand reduction campaigns should be tuned to address the most prevalent themes associated with the product to be targeted. What is needed then, is a way to find themes that consumers associate with a product. To this end, a new product themes discovery algorithm is given next.

2.2.1 Themes discovery algorithm

1. Collect a corpus of tweets from Twitter and/or Facebook postings that contain references to a product of interest for example, *rhino horn*. Say that there are n_s sentences in this corpus.
2. Set m to an integer between 2 and 9. Find a collection of $n_p(m)$ m -grams in this corpus of social media posts as follows. For each sentence in the corpus, randomly generate an *offset*, r_i , $i = 1, \dots, n_s$ where $r_i \in \{0, \dots, m-1\}$. Partition the words in a sentence starting at $r_i + 1$ into m -grams (phrases). Discard the last phrase of a sentence if it contains less than m words. Retain only those phrases that contain at least one common adjective or noun from the list of 500 such words gathered by TalkEnglish (2017).
3. For each $k = 2, \dots, \min\{10, n_p(m)/2\}$, cluster the collection of phrases found in Step 2 with the K -medoids clustering algorithm of Park & Jun (2009) using the dissimilarity metric $d(i, j) = 1 - SIM(ph_i, ph_j)$. Compute the Davies-Bouldin index of cluster separation (Davies & Bouldin (1979)), $DB(m, k)$ for each such clustering solution.
4. Repeat steps 2 and 3 n_{MC} times: each time with newly-generated random offsets. Retain the solution that yields the smallest value of $DB(m, k)$.

5. Using the *indigenous categories* theme identification method (see Ryan & Bernard (2003)), identify a theme for each final cluster by looking for common attitudes, sentiments, or subjects within the cluster's collection of phrases.

Lin & Wu (2009) present an approach to phrase clustering that uses the *K*-means clustering algorithm that is similar to the *K*-medoids algorithm used here. The *K*-medoids algorithm is used here because (a) it operates on inter-phrase dissimilarity alone and hence, cannot produce a cluster centroid that is not an actual phrase (no need to postulate some type of "phrase space"), and (b) it is resistant to outlying phrases.

Used with *k*-medoids clustering,

$$DB(m, k) \equiv \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left\{ \frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right\} \quad (2)$$

where $d(i, j)$ is the dissimilarity between cluster medoid i and cluster medoid j ; and σ_i is the average dissimilarity between cluster l 's medoid and each member of that cluster.

2.2.2 Example

Data in the form of forum posts on the characteristics of the product *ginseng root* was collected from two social media websites on June 27, 2017 (**Table 3**). These sites were found by running a Google™ search on the phrase *ginseng root forums*.

Ginseng root trafficking has similarities with trafficking in other wildlife products such as, for example, rhino horn. Similar to the poaching of rhinos in (mainly) South Africa, ginseng is illegally harvested in the southern U.S.. Similar to Chinese traditions surrounding the medicinal uses of rhino horn, there is a set of traditional medicine beliefs associated with ginseng. And these medicinal beliefs, as with rhino horn, have spawned a large Chinese market for wild American ginseng (Khim 2016).

The product themes discovery algorithm is applied to the above 31-sentence corpus with $m = 7$, and $n_{MC} = 100$. The resulting solution consists of three clusters as shown in **Table 4**. This solution's $DB(7, 3)$ value is 0.573. Experiments (not shown) with m set to values between 2 and 6 produced $DB(m, k)$ values that varied between 0.9 and 5.0.

The indigenous categories method is used to identify a theme for each cluster. To use this method, one notices re-occurring concepts in a set of phrases wherein the speakers define the concept using their words rather than the analyst defining a concept using words the analyst chooses. Here, the first cluster mentions the drug "MDMA" several times, and the drug "straight caffeine." A theme that captures this commonality is *comparisons of ginseng to other drugs*. In the second cluster, mention is made of bodily functions, bodily "systems," one's "immune" system, "beneficial properties," and ginseng being an "effective herb." A theme that summarizes these sentiments is *ginseng's medicinal uses*. In the third cluster, speakers mention ginseng as being "nice social," "in place of cannabis," "me giggling for hours," "loved-up effect," and "feeling lazy in body." A theme that captures these sentiments is *the characteristics of ginseng as a recreational drug*.

2.2.3 Performance evaluation

One way to evaluate a product themes discovery exercise is to assess the stability or distinctness of themes that are recognized across two or more raters (coders). One way to quantify this idea is to compute a measure of inter-rater agreement or reliability. Several such measures exist. One well-known and flexible one is Krippendorff's alpha (see Zapf et al. (2016)). To make this computation, the set of phrase clusters computed by the above algorithm will need to be examined for themes by more than one rater.

3 Ecosystem Data Acquisition

An ecosystem management tool also needs a continuous stream of ecosystem data. As an example of how such data can be automatically acquired, a new method is described for estimating animal abundance from a sequence of remotely-sensed images.

3.1 Remote sensing of megafauna abundance

Due to their size, the abundance of a megafauna species may be estimated with object recognition algorithms applied to satellite images (see Yang et al. (2014)). Consider a region, R that is to be monitored. Say that R is partitioned into a grid of N quadrats. Purchase two sets of images, I_{jk} , $j = 1, 2$ and $k = 1, \dots, n$

Table 3: Forum posts on the topic of ginseng root.

Forum handle	Post
(from http://forums.sherdog.com/threads/ginseng.2558651/)	
Respezy	I ate some fresh ginseng for a while, can't say i noticed a difference.
Shunyata	Ginseng is effective but good ginseng is not cheap.
smart.feller	Having used it, what I noticed was this (results will vary based off the person): Ginseng, as opposed to straight caffeine, seemed to give me a more mild energy source for a more prolonged period. If I had to equate the difference, it'd be like the difference between DMAA (which I would highly recommend for endurance events if you can still get your hands on it) vs. ephedra...both are heavy stimulants, but the immediate body reaction is very different.
Shunyata	Essentially this means that it's really good for strengthening respiratory, digestive, and overall metabolic function. It's excellent for immune health (which is intimately related to the lungs and GI tract which are the major sites of immune activity because the outside world is contacting the inside of the body in these two systems), it is a premium herb for endurance and stamina (coming off of the respiratory and metabolic benefits), and it also has a host of other beneficial properties which tend to fall within those patterns.
(from https://www.dmt-nexus.me/forum/default.aspx?g=posts&t=14876)	
Ginkgo	Today I got some ginseng extract, and I was totally amazed by the effects. I took 1g of the extract that contains 13.5% ginsenosides, twice the recommended dose. It was stimulating, anti-depressive, anxiolytic and actually quite euphoric! I didn't expect that at all, at least not after only one dose. I have seen other people say that high doses of ginseng is a bit like MDMA, and I can certainly see what they mean. I don't agree that it is like MDMA neither in intensity nor the loved-up effect, but it is really, really good! I think I'm going to take ginseng every day for a few weeks and see the results. What's your experiences with ginseng?
rOm	Red Ginseng liquid extract was taken for 10 days recently. It is very potent just shame when there is no more. I don't see it as a mdma sort of thing but always take a good dose but not double or triple. It's a very good superfood.
Ginkgo	I wouldn't call it a superfood, far from it. I eat superfood everyday, maca, goji, blueberries, cacao, hemp seeds, spirulina, etc. No superfood I have tried can even be comparable to ginseng, at least not this extract. I think higher dosages will prove to be a very nice social drug. I agree it's not like MDMA, but some of the qualities are there.
lyserge	Yes I've tried the 25x Ginseng extract from iamshaman and WOW. I tried it on the suggestion of Robert Anton Wilson that ginseng is an effective herb that can be used in place of cannabis for some of the experiments in his books, and was shocked at how effective it was. It had me giggling for hours, just very up and happy. Now that you mention the MDMA connection, and now that I've sampled that family, I can see it's actually closer to the MD* effects than cannabis effects in how stimulating AND euphoric it is. Cannabis leaves me feeling lazy in the body and mind, but ginseng extract is completely stimulating. I gave it to an acupuncturist friend who commented that it contains the Chinese element known as Fire. Iamshaman warns against using it nasally!
jungleheart	Anything that is stimulating makes me crash really badly afterwards. I don't take ginseng, because it makes me feel unbalanced. Maybe occasionally though, because it it such a neat root!

separated in time by one *repeat interval*. For example, the WorldView-4 satellite has a repeat interval of three days (Richardson 2016). Image I_{1k} is colocated with I_{2k} .

The motivation for using two images is that animals differentiate themselves from stationary objects such as terrain features and vegetation by moving to different locations between the two images. This idea has previously been employed in the animal detection algorithms of Oishi & Matsunaga (2014) and Terletzky & Ramsey (2014).

Some type of spatial sampling is usually needed because the range of many megafauna species may be so large that an image of such a large region may be cost-prohibitive. For example, say that rhino abundance in KNP is to be estimated. The area of this region is about 19,485 km^2 . Currently, WorldView-4 image products cost about US\$ 16.00 per km^2 (see the Example, below).

Table 4: Clusters of phrases formed with the product themes discovery algorithm applied to the ginseng root forum data set of Table 3.

Cluster	Cluster Objects (Phrases)
1	<ol style="list-style-type: none"> 1. that at all, at least not after 2. warns against using it nasally! Anything that 3. that it contains Chinese element known as 4. Having used it, what noticed was this 5. though, because it is such neat root! 6. don't agree that it is like MDMA 7. and can certainly see what they mean 8. have seen other people say that high 9. MDMA connection, and now that sampled that 10. it is really, really good! think going 11. but immediate body reaction is very different 12. as opposed to straight caffeine, seemed to 13. agree it's not like MDMA, but some 14. can even be comparable to ginseng, at
2	<ol style="list-style-type: none"> 1. and it also has host of other 2. always take good dose but not double 3. on suggestion of Robert Anton Wilson that 4. It is very potent just shame when 5. of experiments in his books, and was 6. activity because outside world is contacting inside 7. of body in these two systems), it 8. tract which are major sites of immune 9. you can still get your hands on 10. beneficial properties which tend to fall within 11. ginseng is effective herb that can be
3	<ol style="list-style-type: none"> 1. give me more mild energy source for 2. will prove to be very nice social 3. ate some fresh ginseng for while, can't 4. to take ginseng every day for few 5. used in place of cannabis for some 6. good for strengthening respiratory, digestive, and overall 7. me giggling for hours, just very up 8. neither in intensity nor loved-up effect, but 9. leaves me feeling lazy in body and 10. is effective but good ginseng is not

A terrestrial species may occupy a landscape in social groups of a particular size. For example, white rhinos are usually found in groups of 2 to 3 individuals (Shrader & Owen-Smith 2002). This means that such a species will distribute itself across the landscape as a collection of small clusters of individuals. Landscape characteristics such as water and food availability affects where a social group chooses to locate itself in the landscape.

If a population is highly clustered, a sampling unit that has a small area is more efficient than a smaller number of units each with a large area. This is because on average, much of the area of a large unit is wasted in that it will contain no animals. An equivalent case of this characteristic is the use of a ship

to sample a marine animal: a point in the ocean is essentially the (zero) area of the sampling unit. For example, Mier & Picquelle (2008) find that a systematic sample is best for sampling a highly patchy marine animal.

Because satellite imagery is priced by unit of area, to keep imagery costs low, it is desirable to use the minimum amount of area that can provide a useful abundance estimate. Further, unlike ground-based sampling costs, there is no travel-time expense to consider when selecting a sampling design. For example, Panahbehagh (2016) cites reduced travel time as an advantage of adaptive sampling designs. These considerations suggest that a sample composed of satellite images should consist of a large number of small areal extent images located according to a systematic sampling design.

What should be the spatial layout of such a sample? Through a simulation study, Ferguson (1992) finds that a size-30 sample positioned on a region-filling, equally-spaced herringbone layout usually results in estimates with modest standard errors when the total area of animal clusters is about 5% of the region's area.

3.2 Megafauna abundance estimation algorithm

The following *two-image abundance estimation* (TIAE) algorithm uses two co-located images to return an estimate of the abundance of a megafauna species. All images have been converted from their delivered format (usually GeoTIFF) into the ENVI *bands interleaved by pixel* (.bip) format. One way to do this is with the free GDAL utility, `gdal_translate.exe` (GDAL 2017).

1. Let l_a be the maximum length (expressed in numbers-of-pixels) of the animal species for which abundance is to be estimated. Execute the following object-detection procedure first on I_{1k} , and then on I_{2k} to locate the center points of all objects in each image. Set g_k and h_k to the number of these objects in I_{1k} and I_{2k} , respectively.

Object-Detection Procedure: Step across each row of the image. Each pixel's set of spectral values is compared to those of the animal's spectral signature. If a match is found, the pixel's location is added to the set of a previously-started candidate object that shares a corner or side with the pixel as long as doing so doesn't make the object have a maximum inter-pixel distance larger than l_a . Otherwise, a new set of pixel locations is started for a new candidate object. After all pixels have been visited, those candidate objects are retained that have an inter-pixel distance larger than $0.3l_a$.

2. Let c_k be the number of objects that are colocated between these two images. Declare two locations to be the same if their distance apart is less than $l_a/5$.

Compute $g_k = g_k - c_k$ so that g_k becomes the number of object locations in I_{1k} that are empty in I_{2k} . Similarly, compute $h_k = h_k - c_k$ to arrive at the number of object locations in I_{2k} that are empty in I_{1k} .

3. Let $\beta_f > 0$ parameterize the tendency of species members to stay close to their social group. This tendency becomes stronger as β_f becomes larger. For $i = 1, \dots, c_k$, let

$$f_i \equiv \max \left\{ \min_{j \in (1, \dots, g_k)} [d_{1ij}], \min_{j \in (1, \dots, h_k)} [d_{2ij}] \right\} \quad (3)$$

where $d_{ij} \equiv$ distance (*object_i, animal_j*) (in meters) and *animal_j* is the location of the j^{th} unique animal in image l .

Let the probability that a colocated object is actually an animal be

$$\text{logit}(p_{ai}) \equiv \beta_0 + \beta_f \frac{1}{(1 + f_i)}. \quad (4)$$

Using these probabilities, an adjusted count may be computed of colocated animals between the two images that takes into account the possibility that each colocated object is an animal. This adjusted count is $o_k \equiv \sum_{i=1}^{c_k} I_{\{p_{ai} > \pi_a\}}(p_{ai})$ where $\pi_a \in (0, 1)$.

4. The detected number of animals in the k^{th} pair of images is $y_k \equiv (g_k + h_k)/2 + o_k$. The idea behind the first term is to take advantage of the size-2 replication to average out some of the object detection algorithm's false positives and false negatives. The second term adds-in each animal that either did not move between images, or was one of the two animals that occupied the same location between images.
5. The systematic sample's estimate of abundance is $S = N\hat{\mu}$ where $\hat{\mu} = (1/n)\sum_{k=1}^n y_k$. The design-based variance of this estimator is $((N - n)/(Nn(n - 1)))\sum_{k=1}^n (y_k - \hat{\mu})^2$ (see Mier & Picquelle (2008)).

This algorithm has parameters β_o , β_p and π_a that need to be assigned values. The idea behind equations (3) and (4) is that a colocated object has a chance of being an animal if there is a nearby unique animal. Also, no GIS layers on habitat features are needed as the characteristics of how the animals use the landscape is captured by the locations of the g_k and h_k animals in the two images, respectively. And, the colocated animals inherit these characteristics through (4).

3.3 Example

In order to assess the accuracy of the algorithm, it is necessary to count animals in a region where there is a known number of animals. This is a form of validation, called *ground truth* (see Wang et al. 2016). An ideal case would be an image of a protected area containing a population of a large, endangered species. Such images are difficult to obtain due to concerns by protected area managers of the images falling into the hands of poachers.

Therefore, to allow validation of the proposed algorithm, consider the elephant exhibit, “Elephant Valley” at the San Diego Zoo Safari Park (33.09745 latitude, -116.99572 longitude) in Southern California, USA which currently contains 13 elephants (SDZSP 2017). Two images of this park were downloaded in GeoTIFF format from a commercial image provider (TerraServer 2017). Because elephants are relatively large (5–7.5 m long) (TheAnimalFiles 2014), pan-sharpened images (see Tu et al. (2004)) in the “standard +” resolution were purchased at a cost of US\$ 110.00 per image. These images were taken on August 8, 2016 (**Figure 2**), and October 10, 2016, respectively. These dates are separated by more than the satellite’s revisit period. As these particular elephants are captive, however, the TIAE algorithm will not be biased by emigrating or immigrating elephants.

A “standard +” image with area of 6.23 km^2 is 3698 lines (rows of pixels) \times 2773 samples (columns of pixels). Consequently, the inter-pixel distance is

$$\frac{\sqrt{6.23 \times 10^6 / ar}}{2773} = 0.7789 \text{ m} \quad (5)$$

where $ar = 3698/2773$ is the aspect ratio. This resolution is sufficient to detect an elephant.

A 1.5% tolerance is used on each spectral component to decide if a particular pixel’s spectral component matches the corresponding component of an elephant’s spectral signature. This signature is found using Microsoft Paint’s™ Color Picker tool. Here, an elephant’s maximum length is judged to be 6.23 m.

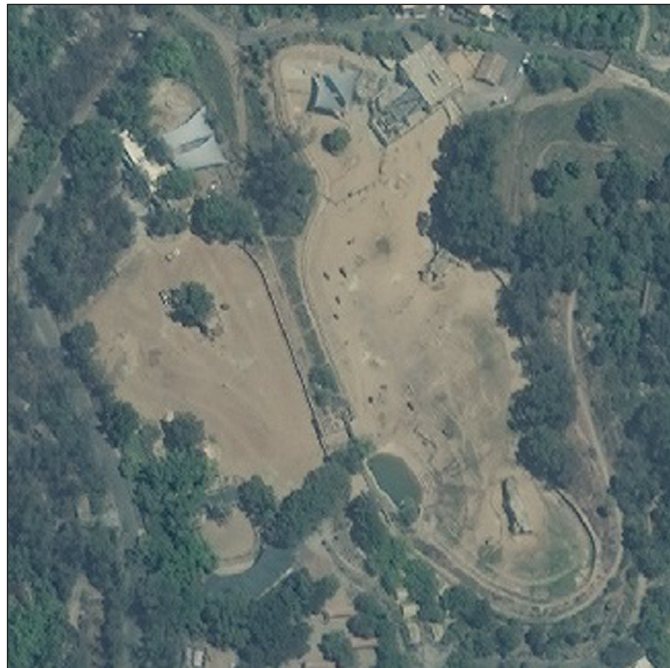


Figure 2: GeoTIFF image of “Elephant Valley” at the San Diego Zoo Safari Park in Southern California, United States on August 8, 2016. When run on this image and an image of the same location photographed October 10, 2016, the TIAE algorithm returns 13 elephants.

Expressing this value in pixels yields $6.23/0.7789 = 8$ pixels. This value is used as an upper limit on the size of objects that are predicted to be elephants.

Use of these constants with the TIAE algorithm results in an elephant abundance estimate of 13 making the algorithm 100% accurate when run on this example's image.

4 An Ecosystem Management Tool's Website

The continuous streams of political-ecological data as exemplified in the previous two sections need to be captured, organized and then distributed to those wishing to use the data to sustainably manage the monitored ecosystem. This data management capability is implemented in the website component of an ecosystem management tool (see Haas (2011, pp. 161–177)). Potential users of an ecosystem management tool include (a) unaffiliated individuals wishing to know the state of particular regions on the planet, (b) federal and state agencies charged with protecting ecosystems, (c) conservation-focused NGOs, and (d) for-profit firms. This last group would use the tool to evaluate the success of profitable *eco-effective* business actions (see Braungart, McDonough & Bollinger (2007)).

The political-ecological data that is intended to support the management of an ecosystem is maintained on a publicly accessible website. This website has the following characteristics:

1. Data contributed to the website is rated as either *sensitive* or *publicly available*. Sensitive data such as the locations of animals belonging to an endangered species, nests of endangered birds, or the locations of rare plants (see for example, Farnsworth (2005), Perinchery (2017); Scheele & Lindenmayer (2017); and Wernick (2017)) are not included under the website's data-download menu.
2. Website menus allow a user to download (a) data on political actions, (b) public data on ecosystem metrics, (c) software used to acquire the data sets and display them, and (d) software that can be used with the website's data to develop ecosystem management policies. All website software is available in source code form.
3. The website is simple and uses neutral language in keeping with its being a source of reliable political-ecological data, and conservation planning software.
4. To maximize portability, only basic HTML tags are used in the website.
5. The website contains an option that allows the entire website to be written to a compressed file and then downloaded by any institution or individual interested in acting as a website mirror. The idea behind this option is that in order to increase the chances that the website is persistent, it should be mirrored across many different hosting institutions.
6. The website is protected from hackers by the host institution's web security infrastructure. This infrastructure is continually updated to reflect the most recent advances in website *intrusion prevention* (see Anwar et al. (2017)).
7. To avoid the costs of computational analyses of website data, computing resources are not offered on the website.

4.1 Operation

The institution that intends to host an ecosystem management tool website first hires a *website coordinator* to perform the following tasks:

1. Maintain the website's availability.
2. Receive new political-ecological data and determine its type (sensitive or publicly available).
3. Enter publicly available data into the website's public data folder, and sensitive data into its hidden, sensitive data folder.
4. Execute protocols for data security, and data release security.

The website coordinator has a fixed two-year tenure and is not eligible for re-appointment.

4.2 Security protocols

As mentioned in Section 1.2, ecosystem data can be used against the ecosystem. One example of this potential for abuse is would-be poachers accessing satellite images that show the locations of individual rhinos.

The potential for abuse begs the question: what security protocols should be followed for keeping an ecosystem management tool's website secure? At first glance, one such protocol should be that locations of animals belonging to an endangered species are not made publicly available. There is a price to be paid,

however, for adhering to such a protocol. There are several scientific reasons for making the original animal locations available to the public. These include the use of such data to develop and statistically estimate the parameters of animal movement models (see Allen & Singh (2016)); studies on animal social behavior (see de Weerd et al. (2015)); and habitat selection studies (see Francis et al. (2017)). One apparent way around this difficulty is to *geographically mask* the image so that animal locations cannot be determined from it (see Zandbergen (2014)). Almost any masking algorithm, however can be reversed given enough time. A safer approach then, is to avoid placing any images that show animal locations in the website's public data folder and instead, implement the following data security, and data release security protocols.

1. A background check is conducted on an individual who is to be hired into a website coordinator position. Once hired, this individual is subjected to *continuous screening* (see Maurer (2016)) in order to detect any event that might cause the individual to become corrupt and begin to sell sensitive ecosystem data to wildlife traffickers. Automated, continuous screening of employees with access to sensitive corporate information is becoming more common. Continuous screening is also known as infinity screening (see Purpura (2013: 136)), post-hire screening, post-employment screening, and post-employment background checks. Some conservation agencies already engage in continuous screening. For example, SANParks conducts random polygraph tests of their rangers as part of their continual monitoring for rangers who might begin to collude with rhino horn traffickers (Presence 2016).

Many firms and universities have staff and protocols for conducting security clearances for university employees who will be working on classified projects for the United States government (e.g. GMU (2017)). Such capability would be leveraged to provide background checks on newly-hired website coordinators and continuous screening of them throughout their two-year tenure. This fixed tenure is seen as needed because doing so gives a website coordinator less time to become corrupt. It is well-known that the size of loss due to fraud increases with the length of tenure of the employee committing the fraud (Kranacher, Riley, Jr. & Wells 2011: 18).
2. When new data is received by the website coordinator, (s)he performs the following tasks.
 - (a) Designates the data as sensitive if it contains specific locations of plants or animals belonging to endangered species.
 - (b) Any electronic files associated with this data are scanned for malware or any other hidden software attachments. Such attacks of course, would be aimed at downloading data that identifies the locations of plants or animals belonging to endangered species.
 - (c) Files deemed to contain sensitive data are collected into a RAR folder (see Library of Congress 2017). This RAR folder is embedded in a *carrier image* via the cryptographic and steganographic security transformation referred to as *subcodsteganography* (Köhler, Lukić & Čik 2017). Then, this security-transformed file along with its *symmetric key* is stored in a hidden folder on the website.
 - (d) If a request is made for the data contained in this image, the website coordinator executes the following newly developed data release security protocol.
 - i. A judgement is made of the likelihood that the requestor plans to use the image for purposes of poaching (either by their own hand or by selling the image to other, would-be poachers). If so, the request is denied. This judgement is based on the outcome of a vote of a panel of three website-affiliated scientists who have reviewed the requestor's credentials. These affiliated scientists hold two-year, non-renewable appointments as data requestor reviewers.
 - ii. If the request is granted, the website coordinator conducts a *video call* (see Martinez & McLaughlin (2017)) interview with the requestor. The purpose of this interview is to receive the *public key* (see Orman (2015: ch. 3)) from the requestor. This key will be used by the website coordinator to encrypt the carrier image's symmetric key before it is emailed to the requestor. The use of a video call interview is intended to defeat any so-called Man-in-the-Middle (MITM) attacks (see de la Hoz et al. (2014)) during public key transmission from the requestor to the website coordinator and works by verifying that the apparent requestor is not in reality an attacker masquerading as the legitimate requestor. In order to do this, the website coordinator compares the face of the person on the video call with known photographs of the requestor, and asks a number of questions to verify that the person in the video call is not an attacker impersonating the requestor. After the website coordinator has decided that the video call interviewee is indeed the legitimate requestor, the requestor is asked to hold up to the video call camera a piece of paper that contains a 256 bit public key expressed as 64 hexadecimal numbers written in four rows of 16 hexadecimal numbers each. The website coordinator takes a screenshot of this piece of paper. The video call

software’s screen-sharing utility is not employed for this activity as that communication channel could be subject to a MITM attack itself. This new, video call-based approach to authentication is an extension of the camera phone approach of McCune, Perrig & Reiter (2005).

- iii. The website coordinator then sends to the requestor the symmetric key of the carrier image that holds the requested sensitive data. This symmetric key is sent as an encrypted email using the requestor’s public key. Once received, the requestor uses his/her public-private key pair to decrypt the image’s symmetric key.
- iv. Finally, the website coordinator emails the subcodstenographically encrypted carrier image to the requestor. The requestor unpacks the sensitive data from the carrier image using the image’s symmetric key received in the previous step.

Conditional on the video call interview successfully detecting a would-be poacher, the remainder of this data release security protocol is essentially secure as it uses cryptographic algorithms that are known to be unbreakable. These two security protocols may appear to be unnecessarily complex. The rationale for this abundance of caution is two-fold. First, cyber crime can be catastrophic because the breaching of a single firewall can expose all data taken on an ecosystem to theft. Second, for the case of sensitive data on the locations of animals and/or plants belonging to endangered species, once a species goes extinct in-part through poaching, it may not be possible to bring it back. In other words, a species is essentially irreplaceable. These two characteristics of biodiversity protection suggest that any online archive of animal location data should be as resistant to intrusion as possible. This view is in agreement with current assessments in the literature towards website security. For a discussion, see Anwar et al. (2017). In essence, then, in lieu of a perfectly secure data center, the website should be run with protocols that are intended to make it as resistant to intrusion and misuse as possible.

4.3 Rhino conservation example

An ecosystem management tool is being developed for the management of white rhinos in South Africa (Haas 2018). The homepage of this tool’s website is displayed in **Figure 3**. The page has an intentionally simple layout.

This website makes available for download data on anthropogenic actions that affect the rhino-hosting ecosystem, estimates of rhino abundance, and a software package (namely, the **id** package) for computing ecosystem management policies. This software includes an implementation of the algorithm for computing the most practical ecosystem management plan. In addition, the website makes available all required input files to run the rhino conservation political-ecological simulator, documentation of this agent-based

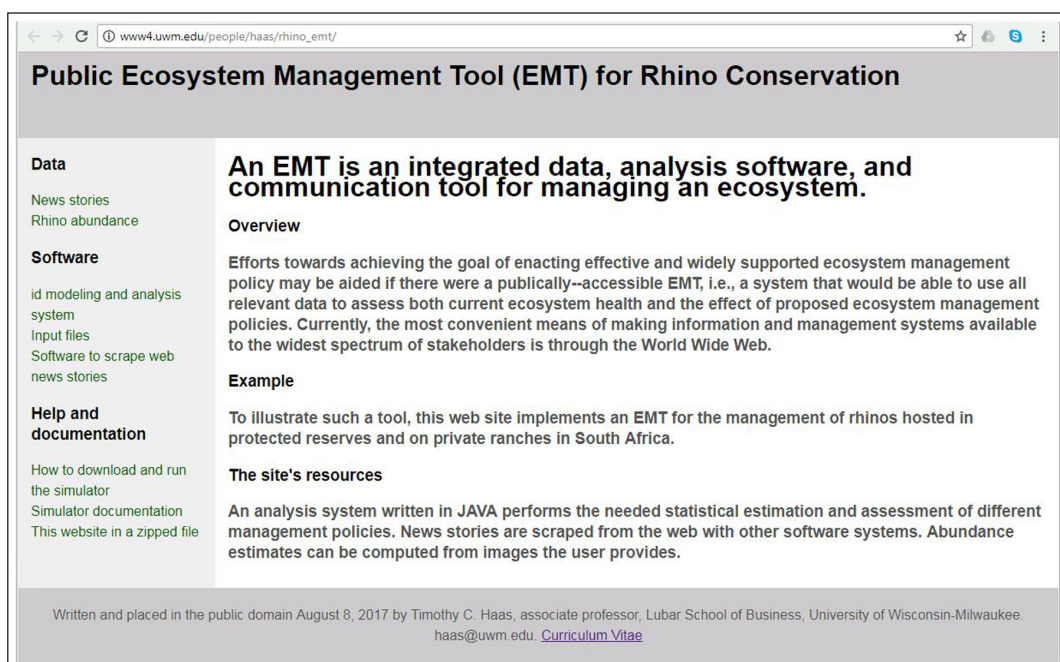


Figure 3: Homepage of the rhino conservation ecosystem management tool.

simulation model, and a user’s guide to running it. Note that the **id** package contains a number of capabilities that are not the focus of this article.

There is a need to build and statistically fit models of political-ecological systems that can be used to assess the effects of proposed ecosystem management options. For example, Haas & Ferreira (2018) develop a political-ecological model composed of an agent-based submodel of poachers interacting with an individual based submodel of the south African rhino population, an agent-based submodel of southeast Asian rhino horn consumers who purchase the poached rhino horn from these poachers, and a submodel antipoaching forces working to curb this poaching activity. Combinations of antipoaching initiatives and economic opportunities for the poachers need to be evaluated as to their probable effectiveness at changing local people’s inclination to poach rhinos and the consequent effect on the rhino population (see Haas & Ferreira 2018). To do this, political-ecological data is needed to statistically validate such a large political-ecological model. One such statistical approach based on *maximum simulated likelihood* (Haas 2011, ch. 4) requires data on some subset of the model’s observable variables. Model validation consists of several tasks including formal statistical parameter estimation, goodness-of-fit assessment, and a sensitivity analysis. When the model is large and stochastic, any of these activities requires large amounts of high-dimensional political-ecological data and access to high performance computing.

To make specific the nature of a submodel’s parameterization, consider the rhino poachers group decision making submodel. This submodel is programmed as a joint probability distribution of discrete random variables defined through tables of conditional probability constants. These constants are the submodel’s parameters. These parameters need to have their values estimated before the submodel can be used to study the effect of different management policies on a poacher’s decision to poach. This *decision making diagram* is shown in **Figure 4**. It is emphasized that this is a computational model that is capable of simulating a group reaching a decision rather than a non-functional, qualitative framework of the decision making process.

Mathematically, parameter values are found that maximize the agreement between the submodel’s decision output and observed decisions made by real-world poachers over a particular time period. Let $\mathbf{out}_i^{(obs)}(t_j)$ be the observed **output action – target** combination of group i at time t_j ; $\mathbf{out}_i^{(opt)}(t_j)$ be the **output action – target** combination computed by the group i ’s influence diagram at that time; and M_{ij} be unity if $\mathbf{out}_i^{(opt)}(t_j) = \mathbf{out}_i^{(obs)}(t_j)$ and zero, otherwise. Then, the overall agreement of all group submodels with the entire set of observations on political actions is $\sum_{i=1}^m g_s^{(i)}(\beta^{(i)})$ where $g_s^{(i)}(\beta) \equiv \sum_{j=1}^T M_{i,j}$.

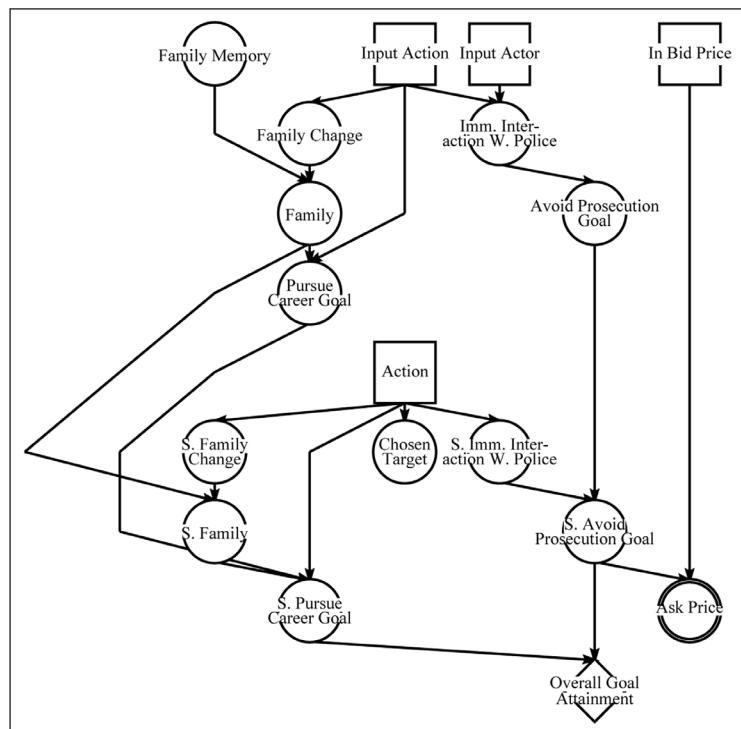


Figure 4: Poachers have two conflicting goals: Pursue Career, and Avoid Prosecution. They have one audience: Family. Square: decision options; Diamond: overall feeling of goal attainment. A decision (action) node along with the Situation Goal nodes affects Scenario Goal nodes. From Haas & Ferreira (2018).

5 Discussion

5.1 Building taxonomies

Building ecosystem-affecting action taxonomies such as the EMAT of Haas (2011: 123–141) is in its infancy. A steady stream of such actions along with an automatic way to discover new ones will allow such taxonomies to be built more quickly and allow comprehensive evaluation of taxonomy quality. In this case, quality refers to the taxonomy's *acceptance, comprehensiveness, completeness, mutual exclusiveness, repeatability, and specificity* (Hansman & Hunt 2005). As the EMAT is an extension of a general political actions taxonomy developed by Leng (1999) and Leng & Singer (1988), it inherits much of this original taxonomy's quality.

5.2 Demand reduction campaigns

Data produced by the new product themes discovery algorithm of Section 2.2.1 is critical to the effective design of advertising campaigns aimed at reducing the demand for wildlife products. For example, expensive advertising campaigns in China that attempt to discourage the consumption of rhino horn assume that the Chinese consume rhino horn as a way to broadcast their success in life, similar to why the Vietnamese consume rhino horn (Truong, Dang, & Hall (2016)). If, in reality, however, Chinese consumers buy rhino horn because they believe it will make them healthy, a campaign focused on social status may not be as effective as expected.

In general, demand reduction campaigns to reduce the demand for wildlife products such as rhino horn need to be tuned to address the beliefs that consumers hold towards these products (Liu et al. 2015). Developing automatic algorithms that can regularly gauge what these beliefs or themes are over many years is crucial to the success of such campaigns as it can take many years to change attitudes. For example, attitudes towards cigarette smoking changed slowly but significantly over the years 1947–1974 in the United States due to anti-smoking campaigns (Warner 1977). Because the consumption of many wildlife products is not physically addictive, it may be easier to reduce demand for them relative to products such as cigarettes or drugs that are.

5.3 Data type interrelationships

A political-ecological system has effects that flow in both directions: anthropogenic actions affect an ecosystem, such as poaching affects rhino population dynamics but the ecosystem, in-turn, affects political and social processes such as wildlife causing crop and livestock damage, and downward trends in rhino abundance causing increased law enforcement activity to attempt to curb the poaching of rhinos.

Therefore, multivariate datasets composed of observations on political actions coupled to ecological actions are needed to formulate, estimate, validate, and then use models of political-ecological systems to help formulate policies that have some chance of making the managed ecosystem sustainable.

As mentioned above, Haas & Ferreira (2016a) build a stochastic, political-ecological model of the rhino management system by coupling together an agent-based submodel of consumer demand for rhino horn, a submodel of rhino poachers, and an individual-based submodel of South African rhino population dynamics. After fitting this model to data, these authors use it to show that changes in demand for rhino horn affect poaching effort which in-turn, affects rhino population dynamics. Further, Haas & Ferreira (2016b) use this model to show that only a coordinated policy has a chance of sustaining the South African rhino population. This coordinated policy would include finding alternative economic opportunities for would-be poachers, increasing antipoaching enforcement efforts, disrupting poaching syndicates (see Haas & Ferreira (2015)), and reducing the demand for rhino horn in countries where it is consumed.

Focusing on the poachers, Haas & Ferreira (2018) use poaching action data to estimate the parameters that represent the decision of a poacher to poach a rhino along with the effect that such poaching has on rhino abundance as captured by the above-mentioned submodel of the South African rhino population.

This overview suggests that some political-ecological data is clearly inter-related such as data on political actions and an ecosystem's responses to those actions. Other types of data, although not directly related, are part of causal processes such as wildlife product themes that point to motivations that drive demand for wildlife products. An early model of a wildlife product consumer's decision making is given in Haas & Ferreira (2016a). This model does not attempt to capture the goals that such a consumer is trying to attain such as social status or health guarantees – but rather focuses only on how the consumer's price sensitivity may affect their decision to purchase rhino horn or not. For this early model, data on the themes surrounding rhino horn would appear to not be related to political actions data or rhino abundance data. But if a consumer decision making model replaced the simple price sensitivity model, the product themes data would become related to these other two data types. It appears then, that the inter-relatedness of different

political-ecological data types and indeed, the very types of data that need to be acquired, depends on what models of the managed political-ecological system are being entertained. More complete models will need more types of data to support their estimation, validation, and use in formulating ecosystem management policies.

5.4 Data release security protocols

It is possible that poachers could purchase satellite images themselves, analyze these images using software taken from (say) the website described herein, and then poach those animals that they detect. Poachers are known to hack into GPS location signals sent to satellites from radio-collared animals (Welz 2017), and to find animals from location information contained in scientific articles that announce the discovery of new species (Neslen 2016). Further, Koh (2015) predicts that poachers will someday locate animals to poach by hacking into the video signal transmitted by monitor drones flown by protected area managers.

An online search performed circa 2018 for articles reporting on the use of satellite imagery by poachers does not produce any results. At present then, poachers may not have the analytical resources needed to post-process satellite imagery. But as image processing software becomes easier to acquire and use, the world may see such use of legally purchased satellite images.

The threat of such a scenario suggests that image providers themselves should employ data release security protocols similar to the ones proposed in Section 4, above. Simple protocols such as only releasing older images may not be practical. For instance, legitimate consumers of these products, e.g. marketing research firms, may require images that are as recent as possible. Hence, a data release security protocol that restricts the image provider to selling images that are of some minimum age may not be followed by the image provider.

Federal and state environmental protection agencies might argue that sensitive ecosystem data should be strictly controlled by them in order to guard against its abuse. But these very agencies may have, unbeknownst to them, corrupt personnel who either give sensitive ecosystem data to wildlife traffickers or join in the slaughter themselves. A case in point is the incident of two SANParks rangers being arrested for rhino poaching (defenceWeb 2017). Leaving ecosystem data security in the hands of designated government agencies then, may not be a 100% secure solution.

A related issue concerning data release security protocols is the issue of acquiring remote imagery for a bespoke date and region. For instance, images used in the elephant abundance example of Section 3.3 are from an image provider's online catalog for which image dates were arbitrarily determined by the image provider. If an ecosystem management tool's website coordinator were to request images taken on other dates, there may be permission issues. This potential difficulty is not explored here but may become problematic if requested dates are very recent.

5.5 Limited access to the web and computing resources

Use of a website to disseminate political-ecological data presupposes unrestricted access to the web. This is not the case in several countries ruled by totalitarian regimes. Providing such information to people in these countries is critically important but this article offers no suggestions for how to do so.

Estimation of simulator parameters, execution of a sensitivity analysis, and computation of a most practical ecosystem management plan requires access to high performance computing systems – although it is possible to execute partial analyses on personal computers. Many interested individuals may not have access to such high performance computing resources and hence would only be able to examine the output of simulator estimation, sensitivity analyses, and most practical ecosystem management plan computations produced by others. But such outputs may not be shareable in the first place. For example, a private firm may wish to keep the results of such analyses confidential to protect its competitive advantage. Therefore, open and planet-wide access to computing resources for purposes of analyzing political-ecological data is needed but is currently a challenge.

5.6 Evaluating an ecosystem management tool

The worth of the proposed ecosystem management tool is in its ability to contribute to biodiversity, i.e., species persistence. According to an ecologically motivated definition given in Haas (2011, p. 10), a species has *low extinction risk* if its probability of becoming extinct at some point in time 50 animal generations into the future is less than 0.01. Therefore, the success of an ecosystem management tool should be measured by computing the managed species' extinction probability 50 animal generations into the future. The tool is more effective as this probability approaches 0.01. To make this calculation, a stochastic simulation model of

the managed political-ecological system needs to be fitted to political-ecological data and then run forward in time under ongoing and proposed management policies to arrive at the species' extinction probability 50 animal generations into the future.

Haas & Ferreira (2018) performed a shortened version of this calculation for the South African rhino population. There, rhino abundance data had been acquired through expensive helicopter surveys and was allowed to be included in a publication only through the existence of a collaborative research project between SANParks, South Africa and an academic. The TIAE algorithm described herein could be used to acquire megafauna abundance data when such research relationships and/or aerial surveys do not exist.

6 Conclusions

The algorithms and examples herein show that political-ecological data needed to support sustainable ecosystem management policies can be automatically acquired at minimal cost. In particular, useful data on ecosystem-affecting political actions can be automatically extracted from the World Wide Web and parsed into a standardized format. Social media posts can be explored with automatic methods to discover themes concerning ecosystem products. Inexpensive high-resolution satellite images can be used to estimate the abundance of megafauna.

This means that certain aspects of ecosystems can be monitored by entities not subject to regime changes in ecosystem-hosting countries or dependent on winning large research grants. Such capability provides the necessary first step towards the sustainable management of at-risk ecosystems by all of the planet's stakeholders – not just those occupying particular governmental positions or members of scientific elites. And, software can be made publicly available for acquiring political-ecological data and using it to explore ecosystem management options.

Before greater citizen involvement in ecosystem management can be achieved, several challenges remain. First, modest but long-term funding streams need to be found to support ecosystem management tool website coordinators – along with appropriate institutions willing to host such websites. Second, permissions need to be secured to regularly purchase remotely-sensed imagery. Third, ecosystem management tool websites need to be accessible to all of the planet's stakeholders. Fourth, these websites need to be designed so that they cannot be used to damage the very ecosystems they have been built to protect. Although the security protocols developed herein constitute an effective first step, this challenge remains a fundamental but open problem in data science. Fifth, although not a focus of this article, high performance computing resources need to be made available to data recipients without regard to their political views or membership in scientific elites.

Competing Interests

The author has no competing interests to declare.

References

- Aarts, B.** 2011. *Oxford Modern English Grammar*, 448. Oxford: Oxford University Press.
- Allen, AM** and **Singh, NJ.** 2016. "Linking movement ecology with wildlife management and conservation." *Frontiers in Ecology and Evolution*, 3. DOI: <https://doi.org/10.3389/fevo.2015.00155>
- American Society for Indexing.** 2018. *Taxonomies & Controlled Vocabularies Special Interest Group*. Available at: www.taxonomies-sig.org/about.htm [Last accessed 7 April 2018].
- Anwar, S, Zain, JM, Zolkipli, MF, Inayat, Z, Khan, S, Anthony, B** and **Chang, V.** 2017. From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. *algorithms*, 10(39). DOI: <https://doi.org/10.3390/a10020039>
- Blei, DM** and **Smyth, P.** 2017. Science and data science. *Proceedings of the National Academy of Sciences of the United States of America*, 114(33): 8689–6892. DOI: <https://doi.org/10.1073/pnas.1702076114>
- Bodin, Ö, Crona, B, Thyresson, M, Golz, A-L** and **Tengö, M.** 2014. Conservation success as a function of good alignment of social and ecological structures and processes. *Conservation Biology*, 28: 1371–1379. DOI: <https://doi.org/10.1111/cobi.12306>
- Braungart, M, McDonough, W** and **Bollinger, A.** 2007. Cradle-to-cradle design: Creating healthy emissions – a strategy for eco-effective product and system design. *Journal of Cleaner Production*, 15: 1337–1348. DOI: <https://doi.org/10.1016/j.jclepro.2006.08.003>
- British Council.** 2017. *Multi-word Verbs*. Available at: <https://learnenglish.britishcouncil.org/en/quick-grammar/multi-word-verbs> [Last accessed 11 September 2017].

- Brombal, D.** 2017. Accuracy of environmental monitoring in China: Exploring the influence of institutional, political and ideological factors. *Sustainability*, 9(324). DOI: <https://doi.org/10.3390/su9030324>
- Ceballos, G, Ehrlich, PR, Barnosky, AD, García, A, Pringle, RM and Palmer, TM.** 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, 1(5): e1400253. DOI: <https://doi.org/10.1126/sciadv.1400253>
- Cedar Creek.** 2018. Data use policies. *Cedar Creek Ecosystem Science Reserve*. Available at: <https://www.cedarcreek.umn.edu/research/datadatause> [Last accessed 3 April 2018].
- Cordeiro, J, Dias, G and Brazdil, P.** 2007. A Metric for paraphrase detection. *Proceedings of the International Multi-Conference on Computing in the Global Information Technology*. IEEE Computer Society. Available at: <http://ieeexplore.ieee.org/document/4137062/> [Last accessed 11 September 2017].
- Daelemans, W, Buchholz, S and Veenstra, J.** 1999. Memory-based shallow parsing. In: *Proceedings of the EACL'99 workshop on Computational Natural Language Learning (CoNLL-99)*, 53–60. Bergen, Norway.
- Davies, DL and Bouldin, DW.** 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2): 224–227. DOI: <https://doi.org/10.1109/TPAMI.1979.4766909>
- defenceWeb.** 2017. A ranger's view on rhino poaching. *defenceWeb*, July 11. Available at: http://www.defenceweb.co.za/index.php?option=com_content&view=article&id=48522 [Last accessed 11 September 2017].
- de la Hoz, E, Paez-Reyes, R, Cochran, G, Marsa-Maestre, I, Moreira-Lemus, JM and Alarcos, B.** 2014. Detecting and defeating advanced man-in-the-middle attacks against TLS. *6th International Conference on Cyber Conflict*, Brangetto, P, Maybaum, M and Stinissen, J (eds.). Available at: http://www.ccdcoe.org/cycon/2014/proceedings/d2r2s2_delahoz.pdf [Last accessed 17 March 2018].
- de Weerd, N, van Langevelde, F, van Oeveren, H, Nolet, BA, Kölzsch, A, Prins, HHT and de Boer, WF.** 2015. Deriving animal behaviour from high-frequency GPS: Tracking cows in open and forested habitat. *PLoS ONE*, 10(6): e0129030. [Last accessed 4 April 2018]. DOI: <https://doi.org/10.1371/journal.pone.0129030>
- EnglishClub.** 2017. *Regular Verbs List*. Available at: www.englishclub.com/vocabulary/regular-verbs-list.htm [Last accessed 11 September 2017].
- Farnsworth, E.** 2005. *Guidelines for Ethical Field Research on Rare Plant Species*. New England Wild Flower Society, January. Available at: <http://www.mass.gov/eea/docs/dfg/nhosp/regulatory-review/nepcpre-searchguide.pdf> [Last accessed 11 September 2017].
- Ferguson, CC.** 1992. The statistical basis for spatial sampling of contaminated land. *Ground Engineering*, 34–38. June.
- Francis, RA, Taylor, JD, Dibble, E, Strickland, B, Petro, VM, Easterwood, C and Wang, G.** 2017. Restricted cross-scale habitat selection by American beavers. *Current Zoology*, 63(6): 703–710. DOI: <https://doi.org/10.1093/cz/zox059>
- Frey, UJ and Cox, M.** 2015. Building a diagnostic ontology of social-ecological systems. *International Journal of the Commons*, 9(2): 595–618. DOI: <https://doi.org/10.18352/ijc.505>
- Gammon Ranges.** 2017. *Gammon Ranges (Nature Conservation Society) Survey*. Data Discovery Portal, Terrestrial Ecosystem Research Network (TERN), The University of Queensland, Australia. Available at: <http://portal.tern.org.au/#/4fb352b1> [Last accessed 2 April 2018].
- Gazzola, M.** 2014. *The Evaluation of Language Regimes*, 375. Amsterdam: John Benjamins Publishing Company.
- GDAL.** 2017. *gdal translate*. Available at: www.gdal.org/gdal_translate.html [Last accessed 9 September 2017].
- GMU.** 2017. *Classified Information and Personnel Security Clearances: University Policy Number 1119*. University Policy, George Mason University. Available at: <https://universitypolicy.gmu.edu/policies/classified-information-and-personnel-security-clearances/> [Last accessed 11 September 2017].
- Guinote, A.** 2017. How power affects people: Activating, wanting, and goal seeking. *Annual Review of Psychology*, 68: 353–381. DOI: <https://doi.org/10.1146/annurev-psych-010416-044153>
- Haas, TC.** 2001. A web-based system for public-private sector collaborative ecosystem management. *Stochastic Environmental Research and Risk Assessment*, 15: 101–131. DOI: <https://doi.org/10.1007/s004770000062>
- Haas, TC.** 2011. *Improving Natural Resource Management: Ecological and Political Models*. Wiley-Blackwell, Chichester, U.K. DOI: <https://doi.org/10.1002/9780470979334>

- Haas, TC.** 2018. *Rhino Ecosystem Management Tool*, online resource. Available at: www4.uwm.edu/people/haas/rhino_emt [Last accessed 21 March 2018].
- Haas, TC and Ferreira, SM.** 2015. Federated databases and actionable intelligence: Using social network analysis to disrupt transnational wildlife trafficking criminal networks. *Security Informatics*, 4(1). DOI: <https://doi.org/10.1186/s13388-015-0018-8>
- Haas, TC and Ferreira, SM.** 2016a. Conservation risks: When will rhinos be extinct? *IEEE Transactions on Cybernetics*, 46(8): 1721–1724. DOI: <https://doi.org/10.1109/TCYB.2015.2470520>
- Haas, TC and Ferreira, SM.** 2016b. Combating rhino horn trafficking: The need to disrupt criminal networks. *PLoS ONE*, 11(11): e0167040. DOI: <https://doi.org/10.1371/journal.pone.0167040>
- Haas, TC and Ferreira, SM.** 2018. Finding politically feasible conservation strategies: The case of wildlife trafficking. *Ecological Applications*, 28(20): 473–494. DOI: <https://doi.org/10.1002/eap.1662>
- Hansman, S and Hut, R.** 2005. A taxonomy of network and computer attacks. *Computers & Security*, 24: 31–43. DOI: <https://doi.org/10.1016/j.cose.2004.06.011>
- Hedley, J.** 2017. *jsoup: Java HTML Parser*. Available at: <https://jsoup.org> [Last accessed 11 September 2017].
- Hoff, E.** 2003. Chapter 7: Language development in childhood. In: *Handbook of Psychology, Vol. 6: Developmental Psychology*, Lerner, RM, Easterbrooks, MA and Mistry, J (eds.). Hoboken, NJ: John Wiley and Sons. DOI: <https://doi.org/10.1002/0471264385.wei0607>
- Ignatow, G and Mihalcea, R.** 2018. *An Introduction to Text Mining*, 310. Los Angeles: Sage Publications.
- Khim, S.** 2016. China's gold rush in the hills of Appalachia. *Foreign Policy*, September 7. Available at: <http://foreignpolicy.com/2016/09/07/the-thrill-of-the-hunt-ginseng-smuggling-poaching-boone-north-carolina-china/> [Last accessed 11 September 2017].
- Koh, LP.** 2015. Poachers expected to use green drones to kill endangered wildlife. *Ecological Society of Australia*. <https://www.ecolsoc.org.au/poachers-expected-use-green-drones-kill-endangered-wildlife> [Last accessed 26 May 2018].
- Köhler, M, Lukić, I and Čik, VK.** 2017. Protecting information with subcodstanography. *Security and Communication Networks*, 13. Article ID 9130683. DOI: <https://doi.org/10.1155/2017/9130683> [Last accessed 15 March 2018].
- Kranacher, M, Riley, RA, Jr. and Wells, JT.** 2011. *Forensic Accounting and Fraud Examination*. New York: Wiley.
- Laurila-Pant, M, Lehikoinen, A, Uusitalo, L and Venesjärvi, R.** 2015. How to value biodiversity in environmental management? *Ecological Indicators*, 55: 1–11. DOI: <https://doi.org/10.1016/j.ecolind.2015.02.034>
- Legalbrief.** 2017. Government slated for withholding rhino poaching data. *Legalbrief*, May 24. Available at: legalbrief.co.za/diary/legalbrief-environmental/story/...withholding...poaching.../pdf/ [Last accessed 11 September 2017].
- Leng, RJ.** 1999. *Behavioral Correlates of War, 1816–1979* (Computer File), 3rd Release, Middlebury College, Middlebury, VT, 1993, Study Number 8606 from the Inter-University Consortium for Political and Social Research (ICPSR), Ann Arbor, Michigan, USA. Available at: www.icpsr.umich.edu [Last accessed 12 January 2010].
- Leng, RJ and Singer, JD.** 1988. Militarized interstate crises: The BCOW typology and its applications. *International Studies Quarterly*, 32: 155–173. DOI: <https://doi.org/10.2307/2600625>
- Lenzen, M, Moran, D, Kanemoto, K, Foran, B, Lobefaro, L and Geschke, A.** 2012. International trade drives biodiversity threats in developing nations. *Nature*, 486: 109–112. June 7.
- Levenshtein, A.** 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8): 707–710.
- Library of Congress.** 2017. *RAR Archive Fiel Format*, Version 5. Available at: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000460.shtml> [Last accessed 19 March 2018].
- Lim, KW and Buntine, W.** 2014. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. *CIKM'14*, November 3–7. [Last accessed 11 September 2017]. DOI: <https://doi.org/10.1145/2661829.2662005>
- Lima, JMT, Valle, D, Moretto, EM, Pulice, SMP, Zuca, NL, Roquetti, DR, Beduschi, LEC, Praia, AS, Okamoto, CPF, da Silva Carvalhaes, VL, Branco, EA, Barbezani, B, Labandera, E, Timpe, K and Kaplan, D.** 2016. A social-ecological database to advance research on infrastructure development impacts in the Brazilian Amazon. *Scientific Data (Nature)*, 3: Article number 160071. DOI: <https://doi.org/10.1038/sdata.2016.71>

- Lin, D** and **Wu, X.** 2009. Phrase clustering for discriminative learning. *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, 1030–1038, 2–7 August. Suntec, Singapore. DOI: <https://doi.org/10.3115/1690219.1690290>
- Liu, Z, Jiang, Z, Fang, H, Li, C** and **Meng, Z.** 2015. “Consumer Behavior” change we believe in: Demanding reduction strategy for endangered wildlife. *Journal of Biodiversity and Endangered Species*, 3(1). [Last accessed 11 September 2017]. DOI: <https://doi.org/10.4172/2332-2543.1000141>
- ILTER.** 2017. Welcome to the LTER network data portal. *Long Term Ecological Research*. Available at: <https://portal.lternet.edu/nis/home.jsp> [Last accessed 3 April 2018].
- Martinez, J** and **McLaughlin, MK.** 2017. The best video conferencing software of 2017. *PC Magazine*, August 30. Available at: <https://www.pcmag.com/article2/0,2817,2388678,00.asp> [Last accessed 19 March 2018].
- Maurer, R.** 2016. Is continuous screening the future normal? *Society for Human Resource Management*, October 28. Available at: <http://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/continuous-screening-posthire-future.aspx> [Last accessed 11 September 2017].
- McCune, JM, Perrig, A** and **Reiter, MK.** 2005. Seeing-is-believing: Using camera phones for human-verifiable authentication. *Proceedings of the 2005 IEEE Symposium on Security and Privacy*. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=1425062> [Last accessed 17 March 2018].
- Michener, WK.** 2015. Ecological data sharing. *Ecological Informatics*, 29: 33–44. DOI: <https://doi.org/10.1016/j.ecoinf.2015.06.010>
- Mier, KL** and **Picquelle, SJ.** 2008. Estimating abundance of spatially aggregated populations: Comparing adaptive sampling with other survey designs. *Canadian Journal of Fisheries and Aquatics*, 65: 176–197. DOI: <https://doi.org/10.1139/f07-138>
- Morgan, TJH, Uomini, NT, Rendell, LE, Chouinard-Thuly, L, Street, SE, Lewis, HM, Cross, CP** and **Evans, C, Kearney, R, de la Torre, I, Whiten, A** and **Laland, KN.** 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, 6, Article number: 6029. DOI: <https://doi.org/10.1038/ncomms7029>
- Neslen, A.** 2016. Poachers using science papers to target newly discovered species. *The Guardian*. Available at: <https://www.theguardian.com/environment/2016/jan/01/poachers-using-science-papers-to-target-newly-discovered-species> [Last access 26 May 2018].
- Newman, G, Chandler, M, Clyde, M, McGreavy, B, Haklay, M, Ballard, H, Gray, S, Scarpino, R, Hauptfeld, R, Mellor, D** and **Gallo, J.** 2017. Leveraging the power of place in citizen science for effective conservation decision making. *Biological Conservation*, 208: 55–64. DOI: <https://doi.org/10.1016/j.biocon.2016.07.019>
- Nuno, A** and **St. John, FAV.** 2015. How to ask sensitive questions in conservation: A review of specialized questioning techniques. *Biological Conservation*, 189: 5–15.
- Oishi, Y** and **Matsunaga, T.** 2014. Support system for surveying moving wild Animals in the snow using aerial remote-sensing images. *International Journal of Remote Sensing*, 35(4): 1374–1394. DOI: <https://doi.org/10.1080/01431161.2013.876516>
- Orman, H.** 2015. Chapter 3: How does secure email work? In: *Encrypted Email*, 100. New York: Springer-Verlag.
- Panahbehagh, B.** 2016. Adaptive rectangular sampling: An easy, incomplete, neighbourhood-free adaptive cluster sampling design. *Survey Methodology*, 42(2): 263–281.
- Park, H** and **Jun, C.** 2009. A simple and fast algorithm for *K*-medoids clustering. *Expert Systems with Applications*, 36: 3336–3341. DOI: <https://doi.org/10.1016/j.eswa.2008.01.039>
- Perinchery, A.** 2017. Tigers, tarantulas and the online hunter. *The Hindu*, July 23. Available at: <http://www.thehindu.com/sci-tech/energy-and-environment/tigers-tarantulas-and-the-online-hunter/article19325605.ece> [Last accessed 19 March 2018].
- Piskorski, J** and **Yangarber, R.** 2013. Information extraction: Past, present, and future. In: Poibeau, T, et al. (eds.), *Multi-source, Multilingual Information Extraction and Summarization 11, Theory and Applications of Natural Language Processing*. Berlin: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-642-28569-1_2
- Presence, C.** 2016. Lie detector tests for KNP rangers. *IOL*, January 29. Available at: <http://www.iol.co.za/news/crime-courts/lie-detector-tests-for-knp-rangers-1977531> [Last accessed 11 September 2017].


- Purpura, PP.** 2013. *Security and Loss Prevention: An Introduction*, Sixth Edition. Waltham, MA: Butterworth-Heinemann.
- Reverso.** 2017. *Irregular Verbs*. Available at: <http://conjugator.reverso.net/conjugation-irregular-ver> [Last accessed 11 September 2017].
- Richardson, D.** 2016. *WorldView-4 Set to Launch After Months Long Delay*. Available at: <http://www.spaceflightinsider.com/organizations/ula/worldview-4-set-to-launch-after-months-long-delay/> [Last accessed 11 September 2017].
- Rissman, AR and Gillon, S.** 2017. Where are ecology and biodiversity in social-ecological systems research? A review of research methods and applied recommendations. *Conservation Letters*, 10(1): 86–93. DOI: <https://doi.org/10.1111/conl.12250>
- Rist, L, Felton, A, Mårald, E, Samuelsson, L, Lundmark, T and Rosvall, O.** 2016. Avoiding the pitfalls of adaptive management implementation in Swedish silviculture. *Ambio*, 45(Suppl. 2): S140–S151. DOI: <https://doi.org/10.1007/s13280-015-0750-9>
- Ryan, GW and Bernard, HR.** 2003. Techniques to identify themes. *Field Methods*, 15(1): 85–109. DOI: <https://doi.org/10.1177/1525822X02239569>
- Scheele, B and Lindenmayer, D.** 2017. Scientists are accidentally helping poachers drive rare species to extinction. *THE CONVERSATION*, May 25. Available at: <https://theconversation.com/scientists-are-accidentally-helping-poachers-drive-rare-species-to-extinction-78342> [Last accessed 19 March 2018].
- Schrodt, PA and Van Brackle, D.** 2013. Automated coding of political event data. *Handbook of Computational Approaches to Counterterrorism*, Subrahmanian, VS (ed.), 23–49. New York: Springer. DOI: https://doi.org/10.1007/978-1-4614-5311-6_2
- SDZSP.** 2017. *Elephant Valley*. Available at: <http://sdzsafaripark.org/park-locations/elephan-valley> [Last accessed 11 September 2017].
- Shrader, AM and Owen-Smith, N.** 2002. The role of companionship in the dispersal of white rhinoceroses (*Ceratotherium simum*). *Behavioral Ecology and Sociobiology*, 52(3): 255–261. DOI: <https://doi.org/10.1007/s00265-002-0506-y>
- TalkEnglish.** 2017. *Top 500 Adjectives*. Available at: <http://www.talkenglish.com/vocabulary/top-500-adjectives.aspx> [Last accessed 11 September 2017].
- Terletzky, P and Ramsey, RD.** 2014. A semi-automated single day image differencing technique to identify animals in aerial imagery. *PLoS ONE*, 9(1): d85239. DOI: <https://doi.org/10.1371/journal.pone.0085239>
- TerraServer.** 2017. *Image Downloads*. Available at: www.terra-server.com/products [Last accessed 11 September 2017].
- TheAnimalFiles.** 2014. *African Elephant*. Available at: http://www.theanimalfiles.com/mammals/elephants/african_elephant.html [Last accessed 11 September 2017].
- Truong, VD, Dang, NVH and Hall, CM.** 2016. “The marketplace management of illegal elixirs: Illicit consumption of rhin horn.” *Consumption Markets and Culture*, 19(4): 353–369. DOI: <https://doi.org/10.1080/10253866.2015.1108915>
- Tu, T, Huang, PS, Hung, C and Chang, C.** 2004. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. *IEEE Geoscience and Remote Sensing Letters*, 1(4): 309–312. DOI: <https://doi.org/10.1109/LGRS.2004.834804>
- Uitto, JI.** 2016. Evaluating the environment as a global public good. *Evolution*, 22(1): 108–115. DOI: <https://doi.org/10.1177/1356389015623135>
- United States.** 2016. Demand reduction strategies to combat illegal trade in CITES-listed species. *Convention on International Trade in Endangered Species of Wild Fauna and Flora*, CoP17 Doc. 18.1.
- United States National Science Foundation.** 2014. *Earth to Data: Making Sense of Environmental Observations*, 14–135. News Release. Available at: www.nsf.gov/news/news_summ.jsp?cntn_id=132973 [Last accessed 5 April 2018].
- Vasani, KA.** 2014. Content evocation using web scraping and semantic illustration. *IOSR Journal of Computer Engineering*, 16(3): 54–60. Ver. IX (May–June). DOI: <https://doi.org/10.9790/0661-16395460>
- Virapongse, A, Brooks, S, Metcalf, EC, Zedalis, M, Gosz, J, Klisky, A and Alessa, L.** 2016. A social-ecological systems approach for environmental management. *Journal of Environmental Management*, 178: 83–91. DOI: <https://doi.org/10.1016/j.jenvman.2016.02.028>
- Wang, S, Li, X, Ge, Y, Jin, R, Ma, M, Liu, Q, Wen, J and Liu, S.** 2016. Validation of regional-scale remote sensing products in China: From site to network. *Remote Sensing*, 8(980). DOI: <https://doi.org/10.3390/rs8120980>

- Warner, KE.** 1977. The effects of the anti-smoking campaign on cigarette consumption. *American Journal of Public Health*, 67(7): 645–650. DOI: <https://doi.org/10.2105/AJPH.67.7.645>
- Welz, A.** 2017. Unnatural surveillance: How online data is putting species at risk. *Yale Environment 360*. Available at: <https://e360.yale.edu/features/unnatural-surveillance-how-online-data-is-putting-species-at-risk> [Last accessed 26 May 2018].
- Wernick, A.** 2017. Online location data on endangered species might be putting them in harm's way. *Public Radio International*, October 15. Available at: <https://www.pri.org/stories/2017-10-15/online-location-data-endangered-species-might-be-putting-them-harm-s-way> [Last accessed 19 March 2018].
- Yang, Z, Wang, T, Sidmore, AK, de Leeuw, J, Said, MY and Freer, J.** 2014. Spotting East African mammals in open savannah from space. *PLoS ONE*, 9(12): e115989. DOI: <https://doi.org/10.1371/journal.pone.0115989>
- Yujian, L and Bo, L.** 2007. A normalized Levenshtein distance metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6): 1091–1095. June. DOI: <https://doi.org/10.1109/TPAMI.2007.1078>
- Zandbergen, PA.** 2014. Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in Medicine*. [Last accessed 5 April 2018]. DOI: <https://doi.org/10.1155/2014/567049>
- Zapf, A, Castell, S, Morawietz, L and Karch, A.** 2016. Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, 16(93). DOI: <https://doi.org/10.1186/s12874-016-0200-9>

How to cite this article: Haas, TC. 2018. Automatic Acquisition and Sustainable Use of Political-Ecological Data. *Data Science Journal*, 17: 17, pp. 1–26, DOI: <https://doi.org/10.5334/dsj-2018-017>

Submitted: 11 September 2017 **Accepted:** 08 June 2018 **Published:** 03 July 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 