Highlights

Sustaining Biodiversity with an Insider-Threat-Resistant Wildlife Cybercrime Investigation Tool

Timothy C. Haas[a,b,c]

- A wildlife crime investigation tool is described that can stop species extinctions.
- This tool is applied to the conservation of the East African cheetah.
- Extinction predictions are integrated with intelligence on wildlife traffickers.
- Those traffickers can be identified who are driving future local extinction events.
- New work makes the tool's criminal intelligence database resistant to insider threats.

Sustaining Biodiversity with an Insider-Threat-Resistant Wildlife Cybercrime Investigation Tool

Timothy C. Haas[a,b,c]

[a] *Lubar College of Business, University of Wisconsin-Milwaukee, 3202 N. Maryland Ave., Milwaukee, 53201, WI, United States*
[b] *UCLA Department of Computer Science, Los Angeles, CA, United States*
[c] *Profitable Biodiversity, Los Angeles, CA, United States*

Abstract

Wildlife trafficking is driving many species to extinction. Most of these transactions are conducted over the internet, social media, and mobile phone networks. Therefore, the investigative abilities of cybercrime analysts could significantly contribute to the preservation of global biodiversity. This article describes a new software tool that can effectively support an international confederation of criminal and cybercrime intelligence analysts in their efforts to curb wildlife trafficking. This tool consists of three modules: A federated, wildlife cybercrime intelligence database system; a political-ecological system simulator; and a social network model of the wildlife trafficking syndicate under investigation. The database module receives predictions of local extinction risks on the species being conserved that have been computed by a credible model of the species-hosting political-ecological system. The confederation integrates these predictions with a social network analysis in order to identify those influential traffickers who are associated with regions having high local extinction risks. This new approach to conducting a wildlife trafficking investigation is illustrated by finding a hypothetical trafficker who is most responsible for the decline of the East African cheetah (*Acinonyx jubatus*) population. To encourage tool adoption, this article also delivers a new solution to the open problem of how to guard a database against insider attacks. This solution is optimized for use in a peer-to-peer, nonhierarchically-

Email address: haas@uwm.edu (Timothy C. Haas[a,b,c])
URL: https://profitablebiodiversity.com (Timothy C. Haas[a,b,c])

managed database system such as the one described herein. A simulation study shows this new insider threat detection algorithm is effective at detecting individuals with access to a secure database who have, unfortunately, gone rogue.

Keywords: wildlife trafficking, extinction risk, cybercrime intelligence, insider threat detection, political-ecological modeling, categorical encoding

---

Abbreviations:

CA: Consistency Analysis
FWCIDBMS: Federated Wildlife Criminal Intelligence Database Management System
GLAD: Global Authorization Derivation
ITD: Insider Threat Detector
IUCN: International Union for the Conservation of Nature
KWS: Kenya Wildlife Service
MNN: Modular Neural Network
NGO: Nongovernmental Organization
SIU: Special Investigation Unit
SQL: Structured Query Language
TAWA: Tanzania Wildlife Management Authority
WCITS: Wildlife Cybercrime Investigation Tool
WTS: Wildlife Trafficking Syndicate

1. Introduction

Wildlife trafficking is driving many species to extinction. Most of these transactions are conducted over the internet, social media, and mobile phone networks (Roy et al., 2024; Wyatt et al., 2022; Demeau et al., 2019; Stringham et al., 2023; Márquez, 2025; Sardari et al., 2026; Chaurasia, 2023). Therefore, the investigative abilities of cybercrime analysts who collaborate with their international counterparts could significantly contribute to the preservation of global biodiversity (Haas, 2023; Sharma et al., 2025). Indeed, to help stem the slaughter driven by wildlife trafficking ECO-SOLVE (2024) calls for

- Enhanced data collection and analysis
- Increased international cooperation

- Improved training and resources for law enforcement
- Better coordination between NGOs and law enforcement

To help the international law enforcement community realize these improvements, this article describes a new software tool that can effectively support an international confederation of criminal and cybercrime intelligence analysts in their efforts to curb wildlife trafficking.

A few definitions are needed as follows. Any individual engaged in the physical acquisition of animals/plants or their parts through the poaching (shooting, trapping, poisoning, digging) of live animals/plants is referred to here as a *poacher*. Poachers, those middlemen who sponsor poaching raids, and those criminals who arrange shipments of poached live animals/plants or their parts are all *traffickers*. These traffickers often belong to a particular *wildlife trafficking syndicate* (WTS) (Haas, 2023). When such a syndicate is modeled as a criminal network using social network theory (Haas and Ferreira, 2015), the criminals who belong to the syndicate are often referred to as *players* (Haas and Ferreira, 2015).

1.1. Using criminal intelligence to curb wildlife trafficking

An international database run by criminal and cybercrime intelligence analysts (hereafter, simply *analysts*) living in different countries is needed to help de-duplicate trafficker identities (Li et al., 2011) and to share intelligence so that these criminals can be put out of business and brought to justice. This may be the only hope for most of the planet's endangered species of flora and fauna (Haas, 2023). One way to form such a shared, international wildlife crime intelligence database is to create a peer-to-peer (P2P) criminal intelligence database that is maintained by a *confederation* of analysts who are employed across several countries. An early form of such a database is developed in Haas (2023). Assume that this confederation has preselected a particular species to be the focus of their wildlife trafficking investigations. The success of this confederation's wildlife trafficking investigations is inversely proportional to the temporally-discounted risk of the preselected species' global extinction (hereafter, *extinction risk*). To succeed then, analysts need to focus their investigations on those suspected traffickers

(hereafter, simply *traffickers*) who are associated with regions that carry the highest local extinction risks of the preselected species. This is because if a species is everywhere locally extinct, it is globally extinct.

To this end, this article describes a *wildlife cybercrime investigation tool* (WCIT). A WCIT consists of three modules: a *federated, wildlife cybercrime intelligence database management system* (FWCIDBMS); a *political-ecological* model and simulator of the system that hosts the preselected species; and a social network model of the WTS built from the intelligence held in the FWCIDBMS

Output from simulation runs of the statistically fitted political-ecological model (hereafter, the *simulator*) includes the local extinction risk for each region within the political-ecological system's spatial extent. The confederation combines this output with a social network analysis of those traffickers associated with regions carrying high extinction risks to produce three lists: A list of those traffickers who should be immediately arrested, called the *Detain list*; a list of those traffickers who should be surveiled, called the *Surveil list*; and a list of those near-future trafficking actions that should be interdicted, called the *Interdict list* (Haas, 2023).

The WCIT described herein is new.

1.2. Article deliverables and layout

This article delivers the following.

1. A tool for investigating the trafficking of a preselected species that is informed by a *credible* model of the political-ecological system that hosts that species
2. An FWCIDBMS that is a synthesis and extension of both the political-ecological database described in Haas (2021) and the federated wildlife cybercrime intelligence database described in Haas (2023)
3. An automatic procedure embedded in the FWCIDBMS for detecting insider attacks against it
4. A demonstration of the WCIT being used to help conserve the East African cheetah (*Acinonyx jubatus*) population

See Haas (2020), and Haas (2024a) for the definition of *credible* that is used in this article.

The FWCIDBMS is described in Section 2. Section 3 contains a description of its built-in procedure for detecting insider threats. The use of simulator output and social network analysis to identify traffickers to detain or surveil along with WTS actions to interdict is detailed in Section 4. As an example, the WCIT is applied in Section 5 to the conservation of East African cheetah. Issues surrounding this tool are discussed in Section 6. Conclusions are drawn in Section 7.

The locations of the freely available WCIT – namely, the associated data, examples, JAVA^{TM} source code, and scripts are given in the Data Availability section, below.

2. The FWCIDBMS

2.1. Nomenclature

A *relational* database consists of observations (records) on *entities*. An entity is a particular object or event in the real world. These entities are characterized by their *attributes* (Haas, 2023). A *query* against a database is a request from a user to add to the database, delete from the database, or copy from the database – a set of records concerning selected entities and the values of selected entity-specific attributes. A query produces a *query result*. This result can contain many records of entity attribute values. Here, a bundle of query results where each query result is generated from a separate query, is referred to as a *set of query results* rather than simply *query results* in order to sharply distinguish a collection of query results from the (possibly) many records inside one particular query result.

2.2. Database entities

An FWCIDBMS consists of both open access data and secure intelligence acquired by confederation members (hereafter, simply *members*) during the course of their investigation and surveillance operations. Extending the criminal intelligence database developed in Haas and Ferreira (2015) and Haas (2023), database entities are `Traffickers`, `Phone Calls`, `Vehicles`, `Firearms`, `Bank Accounts`, `Wire Transfers`, `Wildlife Product Shipments`, and `Arrests`. Shipments can consist of live animals/plants or their parts, e.g. tiger bones.

Attributes of these entities are listed in Table 1.

Entity	Attributes	Scale
Trafficker	name	nominal
	town	"
	country	"
Phone	owner	"
	phone number	"
Link	from-trafficker	"
	to-trafficker	"
Vehicle	owner	"
	registration number	"
Firearm	owner	"
	serial number	"
Bank account	owner	"
	account number	"
Wire transfer	originator	"
	receiver	"
	amount	continuous
Wildlife product	origin	nominal
shipment	destination	"
	product type	"
	size	continuous
Arrest	trafficker	nominal
	date	continuous
	arresting authority	nominal

Table 1: Entities and their attributes contained in the FWCIDBMS. Most of these attributes are nominally-valued.

2.3. Queries

Typical queries against this database would include:

1. Phone calls wherein the call's transcript contains the word "poach"
2. Calls wherein the transcript contains the word "price, deal, animal/plant part"
3. The where, when, and size of wildlife shipments over a designated time interval.

4. Trafficker arrests: date, charges and what was seized, e.g. wildlife products, phones, firearms, and/or vehicles.

2.4. Database access privileges

The *logistics node* of the FWCIDBMS (Haas, 2023) stores each member's contact information along with auditing and security information on each database node. This node also manages membership dues, and controls database access with the Global Authorization Derivation (GLAD) protocol (Castano et al., 1997). The use of the GLAD protocol within the FWCIDBMS ensures that a single member's security concerns are not dismissed by a cadre of other members (Haas, 2023).

3. Protecting a Database from Insider Threats

Members are all peers. This means that members in one country cannot force members in other countries to share their criminal intelligence by uploading it to the FWCIDBMS. Conversely, once uploaded, each member can only *trust* other members to (a) not access their uploaded data for malicious purposes and/or (b) not damage such data. Hence, such a database management policy begs the question: Why should one analyst in one country, trust another who lives in some other country and for whom they know nothing about? Indeed, a member might be bribed or blackmailed into using their database access privileges to *attack* the criminal intelligence database itself. Such attacks can take many forms including (a) the downloading of data with the intent to distribute it to the very criminals the confederation is gathering evidence on, or (b) the editing of database entries with the intention of undermining investigations. These member-attack potentialities are called *insider threats* (Kul et al., 2020).

Zero trust database access safeguards (Wang et al., 2025) do not apply to this case because members already possess GLAD-determined database access privileges.

Agencies within the United States government have a similar problem: How to enable the sharing of military/terrorism intelligence with intelligence agencies in other countries? These foreign intelligence specialists do not work for the United States and are under no compulsion to cooperate with United

States intelligence specialists. For military/terrorism intelligence systems, in particular, insider attacks can cause serious damage. Recent examples include then-president Trump's sharing of Israeli intelligence with the Russians (Gramer, 2017), and the attacks carried out by Edward Snowden, Chelsea Manning, and Nghia Hoang Pho (Raywood, 2018).

Detecting those insiders who launch these attacks is challenging within a hierarchically controlled database running under a *role based access control* (RBAC) policy (Marquis, 2024). In RBAC, each member is assigned a particular role that has associated with it, a fixed set of database access privileges. But in the federated wildlife cybercrime database developed in Haas (2023), all members have the same role and their database access privileges are automatically controlled via GLAD.

As highlighted in Haas (2023), this lack of differentiated roles means that an international wildlife cybercrime intelligence database requires different cybersecurity solutions than those typically installed in corporate or governmental databases wherein database access is determined and enforced through a strict hierarchy of organizational authority. To address this vulnerability of a role-free database, this article describes a new and tested algorithm for detecting insider threats called here, the *insider threat detector* (ITD). The ITD makes an FWCIDBMS resistant to attacks by its own members.

The ITD is described and evaluated below.

3.1. Challenges and previous work

A confederation's FWCIDBMS, being shared by peer analysts who reside in many different countries, can be vulnerable to insider attacks. This vulnerability will be recognized by any analyst who might be thinking of becoming a member of such a confederation. Therefore, the database needs to have in it, a system for safeguarding itself from such attacks. Because of the nonhierarchical control structure of this database and the voluntary nature of becoming a confederation member, such safeguards need to be convincing to both current and potential members. Otherwise, those very analysts who could contribute the most to the fight against wildlife trafficking will hesitate to join the confederation since by doing so, they may see their highly

confidential criminal intelligence stolen or corrupted by any number of rogue members. In other words, a confederation will only be effective at curbing wildlife trafficking if its FWCIDBMS is running an insider threat detection technology that is capable of detecting members who have gone rogue.

Haas (2023) offers one way to manage a federated P2P database but offers only a member-initiated way to detect whether some other member is an insider threat. Once a member claims another member is an insider threat, Haas can only offer a voting-based way to corroborate this member's accusation (Haas, 2023).

In contrast to this member-initiated approach, modern insider threat detection algorithms watch a member's pattern of queries or query results and when these patterns change, this member is declared by the algorithm to be a threat. These *within-database* methods of detecting insider attacks can be made part of the automatic functioning of an FWCIDBMS.

3.2. ITD algorithm

Here, insider threats are detected using a modified form of the *data-centric* method developed in Mathew et al. (2010). Specifically, a *modular neural network* (MNN) classifier (Anand et al., 1995) is used to predict whether member j's query results are anomalous or not. If they are declared to be anomalous, this is taken as evidence that member j is using their access to the confederation's database for malicious purposes.

In the FWCIDBMS, each time member j sends a query to the database, the ITD uses a trained MNN to predict the query's author. This is done by presenting the query result's *S vector* (Mathew et al., 2010) to the MNN and receiving back, a prediction of the query's author. This is done for the results from two consecutive queries sent by member j to the database. If, on both results, an anomalous result is declared by the algorithm, member j is flagged by the ITD as an insider threat. An anomalous query result is declared if (a) the MNN does not predict member j to be the author of the query, and (b) $z_{norm}/z_{average} > p$, where z_{norm} is the value of member j's MNN output node, and $z_{average}$ is the average of these values over all of member j's query results in the MNN's training data set. The value of $p \in (0,\ 1)$ is a threshold parameter.

As opposed to examining only one query result as is done in Mathew et al. (2010), the ITD waits until two consecutive query results are available before testing for an anomaly. In other words, the algorithm waits for a temporal pattern to emerge before testing. This aspect of the algorithm makes it particularly sensitive to *low and slow exfiltration* attacks (see below). The requirement that two conditions hold before a query result is declared to be anomalous helps to guard against an excessive number of false positives. A false positive can unnecessarily disrupt an investigation while the confederation determines whether the member in question has actually gone rogue or not.

In the example below, the ITD's MNN employs two hidden nodes per output node. Specifically, output node i represents the prediction that, given the input vector's values, the query was authored by member i.

3.3. Training the ITD

A set of at least two query results from each of the confederation's r members is used for train the MNN where the number of the MNN's modules is set to r. This means that each neural network module will become a specialist at recognizing query results from its assigned member. The Quasi-Newton-based algorithm of Setiono and Hui (1995) is used to fit the MNN's parameters to the training data set. The objective function consists of averaged prediction error rates for each member. Doing so eliminates the data imbalance problem wherein some members submit fewer queries to the database than other members.

To allow for gradual changes in database content through time along with changes in what is pulled from the database through queries as investigations proceed through time, only the most recent two, non-anomalous query results generated by each member are used to train the MNN. MNN retraining occurs as soon as half of the members have generated at least two non-anomalous query results since the last retraining activity.

A modified S vector is used to form summary measures on the attribute values contained in a query result. For this to work, each attribute on each entity in a query result needs to be uniquely identified because attributes are specific to the entity that they describe (Table 1).

Let n_d be an entity's total count (tally) in the database. Let $n_i^{(m)}$ be the tally of this entity in the i^{th} query result. In the database, arbitrarily index the unique values of a particular discretely-valued attribute owned by a particular entity with the integers, $\{1,\ 2,\ \ldots,\ n_{att}\}$. This is called *ordinal encoding* of the categorical attribute (Kosaraju et al., 2023). Several entities may have the same value of an attribute. In this case, attribute values may repeat in a query result. For instance, the attribute `country_of_residence` owned by the `trafficker` entity might have the unique values of Kenya, Tanzania, South Africa, Nambia, and Mozambique. These labels might be given the arbitrary index values of 1, 2, 3, 4, and 5, respectively. This yields $n_{att} = 5$. A query result might contain $n_i^{(m)} = 7$ occurrences of the `trafficker` entity with `country_of_residence` attribute values of: {South Africa, Tanzania, South Africa, Kenya, Mozambique, Mozambique, Mozambique}. The index values in the query result would therefore be {3, 2, 3, 1, 5, 5, 5} or, in sorted order, {1, 2, 3, 3, 5, 5, 5}. It is emphasized that n_{att} is unrelated to n_d. For instance, in the above example, the number of trafficker entities in the database might be large, e.g. $n_d = 275$.

Attributes may be nominally-, ordinally-, or continuously-valued. The next section shows how each of these scales is transformed into variables that are suitable for inclusion in the MNN's input vector.

3.4. Summary statistics for attribute data

As described above, if an attribute is nominally-valued, an arbitrarily assigned but unique index value is given to each of its labels as it is initially read into the database. These index values allow a nominally-valued attribute to have S vector measures assigned in exactly the same way as a true, ordinally-valued attribute such as the number of vehicles owned by a trafficker. Therefore, it suffices to consider only an ordinally-valued attribute.

Say that such an attribute has n_{att} unique labels in the database. For instance, the attribute, `name` of a `trafficker` entity will have almost as many unique values as there are unique individual trafficker entities in the database. The histogram of such a nominally-valued attribute in a query result would most likely have exactly one observation in each histogram bin when a bin is defined to be a single, unique attribute value. For such an

attribute, a histogram does not convey useful information about the nature of a query result. A criminal intelligence database, however, focuses on individuals and their characteristics. Such databases will typically contain many entity-attribute relationships wherein each entity possesses a unique label of a nominally-valued attribute (Table 1).

Therefore, instead of histogram-based measures, the normalized sample median, normalized sample interquartile range (IQR), and the relative tally of attribute values are used to summarize the ordered index values of an ordinally-valued attribute in a query result. The idea is to use the sample median to locate the attribute, the IQR to measure its dispersion, and the tally to measure its magnitude. How these measures are computed is described next.

Ordinal encoding (also known as *label encoding*) is one of several *categorical encoding* methods (Kosaraju et al., 2023) that are widely used in machine learning to transform a categorical variable into a variable that is more amenable to machine learning algorithms. Because of the potential for a large number of neural network input variables such as the input vector used by the ITD's MNN, finding a low-dimensional representation of a categorical variable is particularly important when a neural network is to be trained on categorical data (Hancock and Khoshgoftaar, 2020). Care needs to be exercised however, to ensure that the artificial ranking of labels produced by ordinal ranking does not compromise the neural network's ability to predict values of its output variable (`query author`).

The *quantile function*, $Q(p)$ is the generalized inverse of the *cumulative distribution function*, $F(x)$: $Q(p) = \inf\{x : F(x) \geq p\},\ 0 < p < 1$ (Redivo et al., 2023). The quantile function may be estimated in two steps. First, assign an index value to each unique attribute value contained in the database, e.g. assign a unique integer value to each unique string of the `trafficker` entity's `name` attribute. Let $x_1 < x_2 < \ \ldots \ < x_n$ be these ordered attribute index values. Second, employ a well-known quantile estimator to estimate $Q(p)$ from these ordered index values. This quantile estimator works as follows.

Letting $\lfloor . \rfloor$ be the *floor* function, if $2\lfloor np/2 \rfloor < np$, then np is not an integer. In this case, let $\hat{Q}(p) = x_{\lfloor np \rfloor + 1}$. If np is an integer, let $\hat{Q}(p) = x_{np}$ (SAS, 2025). As an illustration of how this estimator works, consider the

sample {1, 2, 3, 4} taken from the ordinal random variable, X. Say that each of its four possible values are equally likely. The sample quantile, $\hat{Q}(0.2) = 1$ because $F(1) = 0.25 > 0.2$ – satisfying the above quantile function definition. And for the sample, {1, 2, 3, 4, 5} taken from the ordinal random variable, Y having five equally likely values, $\hat{Q}(0.7) = \hat{Q}(0.8) = 4$.

This quantile estimator is used to compute the sample median, $\hat{Q}(0.5)$ and the sample IQR, $\widehat{IQR} = \hat{Q}(0.75) - \hat{Q}(0.25)$ from the three sample *quartiles;* $\hat{Q}(0.25)$, $\hat{Q}(0.50)$, and $\hat{Q}(0.75)$. For example, say that a query result contains a set of names: {Ben, Jerry, Ralph, Linda, Mary}. Sorting the associated set of index values might yield: {1, 4, 9, 17, 23}. This data set's sample median and sample IQR are 9 and $17 - 4 = 13$, respectively. The third statistic used to summarize the values on a particular attribute in the i^{th} query result is the tally of these values, $n_i^{(m)}$. In this example, $n_i^{(m)} = 5$.

Because all variables in the MNN's input vector need to take values on the unit interval, the median and IQR are divided by n_{att}, and the size measure, $n_i^{(m)}$ by n_d before they are added to the MNN's training data set.

Unlike a nominally- or ordinally-valued attribute, the values of a continuously-valued attribute such as the monetary size of a wire transfer, are used directly to summarize the query result rather than their associated index values. But similar to a nominally- or ordinally-valued attribute, the three statistics used to summarize a continuously-valued attribute in the i^{th} query result are the sample median divided by att_{max}, the sample IQR divided by att_{max}, and $n_i^{(m)}$ divided by n_d where att_{max} is the maximum value of the attribute in the database.

3.5. Evaluating the ITD's performance

3.5.1. Types of insider attacks

Advance Datasec (2026) gives a taxonomy of the main strategies that insiders use to attack a database. One of the most difficult attack strategies to detect is low and slow exfiltration. This attack involves the extraction of small amounts of data over several queries in order to evade monitoring tools that are tuned to detect sudden and large changes in the insider's query results.

An *adaptive insider attempting to evade detection* may employ several different attack types that often include exfiltration (Inayat et al., 2024).

Therefore, for purposes of this article, an *adaptive insider attempting to evade detection* is considered an instance of low and slow exfiltration. A *data breach* is the unauthorized or illegal disclosure of an organization's confidential information and hence is an end-goal of low and slow exfiltration rather than an attack type itself.

So called *mimicry attacks* (Goyal et al., 2023) involve an outsider gaining database access credentials of an insider and then behaving in a similar manner to the insider (mimicking) to avoid detection all the while performing low and slow exfiltration to extract data for malicious purposes. This type of attack is similar to the low and slow exfiltration attack by an insider. Because the present article is focused on members that may become threats rather than outsiders gaining access to the FWCIDBMS, this type of attack will not be pursued further here.

Many databases can only be queried with queries written in the Structured Query Language (SQL). Paul et al. (2024) give a comprehensive review of the many types of *SQL injection attacks* (SQLi). These types of attacks are also referred to as *stealthy querying* (Arif and Wani, 2025). These authors develop a system to defend against six frequently-observed attack types: *Benign Data, Authentication Bypass, Blind SQL Injection, In-band SQL Injection, Remote Code Execution, and Denial of Service.* These attack types are engineered by malicious outsiders to break into a secure database and either steal data or corrupt the database itself. If these outsiders already had unrestricted access to the database, they would not bother with such attack types but rather, would directly extract or corrupt the database by issuing complex queries against the database.

The type of threat that might keep an analyst from joining a federated criminal intelligence database is where a member (an insider) decides for a variety of reasons to attack the database using his/her GLAD-determined access to that database. Such an attack-type is not the focus of SQLi attacks. Therefore, this article will not consider SQLi attacks further.

3.5.2. Data sets for evaluating cyber attacks

Muñiz et al. (2026) develop a synthetic data set called SPEDIA designed to test the effectiveness of insider threat detection tools. These authors also review several other such synthetic data sets. The most well-known is the

CERT Insider Threat Test Dataset Lindauer (2020). These data sets do not contain SQL query result data. But this is the only information that the ITD uses to detect insider threats. A literature search failed to identify a data set composed of SQL query results built for the purpose of testing query result-based insider threat detection tools.

3.5.3. Simulating query results

Because of this lack of published query result data sets, a data set has been created that includes simulated query results generated by a rogue insider who is conducting a low and slow exfiltration attack against a database. This type of attack was chosen because it would be the attack of choice if a member were bribed/compromised to provide data on some of the traffickers contained the confederation's database and would be attractive to that member because it is notoriously difficult to detect.

A JAVA program has been written by the author to simulate attribute data at the level of an ordered index. When $n_j^{(m)} > 1$, let s_i be the random interval between sampled index values where, for member j, $n_j^{(max)}$ and d are given integer constants. Let the starting index value, γ be randomly chosen from the integers $1, \ldots, n_{entities}/d$ where $n_{entities}$ is the number of entities in the database that possess a value on the attribute.. Let the maximum interval size be $\beta = (n_{entities} - \gamma)/n_j^{(m)}$, $\alpha_j = \min(n_j^{(max)},\ \beta)$, and

$$s_i \sim \text{Discrete-Uniform}\,(1,\ \alpha_j)\,. \tag{1}$$

The value of $n_j^{(max)}$ controls the disperson of the simulated index values and takes on one of the values in the set $\{1, \ldots,\ \lfloor n_d/(n_j^{(m)} - 1)\rfloor\}$. The chosen value for $n_j^{(max)}$ also indirectly affects the median of the sampled index values.

For instance, if $n_d = 400$ and $n_j^{(m)} = 51$, then $n_j^{(max)}$ could be one of the integers one through eight. A simulated query result on this attribute is $\{x_1,\ \ldots,\ x_{n_j^{(m)}}\}$ where $x_1 = s_1$, and $x_i = x_{i-1} + s_i$, $i = 2, \ldots,\ n_j^{(m)}$.

When $n_d = 400$ and $n_j^{(m)} = 50$, a query result generated by member j might be characterized by an $n_j^{(max)}$ value of three. At some point in the future, however, this same member might, through bribery, extract a different query result on the same attribute. Such an insider attack could be simulated by setting $n_j^{(max)}$ to the value seven.

This simulation algorithm is new.

3.5.4. Clarifying example

Consider a confederation consisting of two members. Say that each member has their own, unique set of SQL "`where`" conditions when submitting queries for the values on two attributes: `trafficker-name`, and associated `vehicle-registration-number`. Finally, say that there are $n_d = 300$ unique values on each of these attributes. Member 1's queries result in $n_1^{(m)} = 5$ entities, and member 2's queries result in $n_2^{(m)} = 20$ entities. This behavior is simulated by generating a data set consisting of a size-40 set of query results on member 1 and a size-40 set of query results on member 2. These simulated sets of query results are generated with $n_1^{(max)}$ set to two, and $n_2^{(max)}$ set to 15.

Using one hidden node for each member, six input variables, and 18 parameters, the ITD fitted this data set in 7,367 function evaluations. The objective function's value at a randomly-generated starting point was 1.0014, and 2.54E-14 at convergence. This MNN configuration applies only to this clarifying example.

Then, one day, member 2 queries the database for these same two attributes but now restricts the query to those traffickers who have bank accounts. Traffickers in this new query result will be somewhat different than those that are usually returned to this member. This behavior is simulated with member 2's new query result given in Figure 1. The ITD evaluated this new query result from member 2 and declared member 2 to be an insider threat.

Attribute 1	Attribute 2
2	4
4	10
7	12
11	22
18	23

Figure 1: Member 2's new query result in the clarifying example.

3.6. Test data set

A large test data set is simulated that represents a plausible scenario of how two rogue members could practice a low and slow exfiltration attack

against a confederation's wildlife crime intelligence database. This data set is used below to evaluate the ITD's performance.

This test data set has the following characteristics.

1. The database contains intelligence on 1000 traffickers.
2. There are $r = 10$ confederation members. Of these, only two have gone rogue.
3. Each member, whether rogue or trustworthy, generates five query results over a year. Trustworthy members generate query results that each contain $n_i^{(m)} = 60$ entity-attribute records. Rogue members also generate results of size 60 at each of the first three time points.
4. Each of these query results contains values on two nominally-valued attributes. Each attribute has 1000 unique categories.
5. Trustworthy members produce query results wherein $n^{(max)} = 4 + m$ and $d = 10 + 2m$ where $m =$ member ID number: $m = 1, \ldots 10$.
6. But starting with the fourth query, in an effort to avoid threat-detection monitoring systems that are set to trigger on large increases in result-size from one query result to the next, the rogue members issue queries against the database that generate result-sizes that are not extremely different than those they generated at time points 1, 2, and 3 – nor extremely different than their colleagues. Specifically, starting on the fourth query, Rogue member #1 generates 80 records per query result, and rogue member #2 generates 90 per result.

   These two patterns of query result-sizes is one way to simulate a low and slow exfiltration attack by each of these rogue members against an FWCIDBMS.
7. Further, starting with their fourth query, these rogue members start pulling attribute values from different entities (traffickers). These different entities possess different attribute values than those contained in their first three query results. These rogue members do this because they are intent on gaining information on traffickers who are not currently the subject of confederation investigations. These sets of different attribute values are simulated by setting $n^{(max)}$ to 20, and 30 for rogue members #1 and #2, respectively. Further, d is set to 2, and 3 for these two members, respectively.

3.7. Error rates and comparison with a KNN-based algorithm

Using this test data set, the ITD is compared to a simpler alternative used to detect insider threats, namely a *K-Nearest Neighbors* (KNN)-based algorithm (Al-Shehari et al., 2024; Bao and Gao, 2025). The KNN-based algorithm uses the previous (time point = 3) set of query results to find the query result that is closest to a member's query result at time point 4. KNN employs only one nearest neighbor and computes euclidean distance using the same three summary measures as used by the ITD. The author of this closest query result is the KNN-based algorithm's prediction of the author of that query result. If this author is different from the actual author, an anomaly is declared. Note that a KNN classifier does not need to be trained. Also note that the KNN-based algorithm checks for an anomaly using only one query result rather than waiting for two consecutive query results to occur as is done by the ITD.

Several performance measures are computed: The number of false positives (FP), number of false negatives (FN), number of true positives (TP), number of true negatives (TN), and the F1 score. The mathematical definition of this score is

$$\text{F1 score} = \frac{2PR}{P+R} \tag{2}$$

where $P = TP/(TP + FP)$ (precision), and $R = TP/(TP + FN)$ (recall) (Hand et al., 2021). Also computed are the detector's accuracy: (TP + TN) divided by the number of classifications; and the false positive rate: FP divided by the number of classifications.

The synthetic data set generated above is run through both the ITD and the KNN-based algorithm. The results show that the ITD is more accurate and has a lower false positive rate (FPR) than the KNN-based algorithm (Table 2).

Measure	ITD	KNN
FP	0	4
FN	1	1
TP	1	1
TN	8	4
F1 score	0.667	0.286
Accuracy	0.9	0.5
FPR	0.0	0.4

Table 2: Performance measures of two insider threat detection algorithms: the ITD, and a KNN-based algorithm. Ten members have trusted access to the FWCIDBMS.

The value p in the ITD algorithm can be varied from 0.07 to 0.04 with no change in ITD's performance (Table 2).

Runtime for this performance assessment that included fitting of the MNN is 0.26 seconds. Such speed suggests that the ITD could be inexpensively scaled to support a confederation of 100 members – both in terms of the increased computer time needed for ITD training, and for the increased number of ITD runs required for the increased number of queries collectively issued by these members. Such a large confederation would have the investigative capacity to pursue most active wildlife traffickers. Note that scaling the ITD is in terms of the number of members – not the number of traffickers. Although many individuals have the skills to become traffickers, a much smaller number have the credentials to become analysts.

3.8. Integrating insider threat detection into the FWCIDBMS

Because the federated database of Haas (2023) uses PowerShell™ scripts to coordinate its various operations, the inclusion of the ITD can be incorporated within a computationally-efficient language (such as JAVA) that runs outside of the relational database software package. Specifics of how this integration is accomplished follow.

The logistics node is extended in the FWCIDBMS to automatically collect a copy of each query result that is generated from every query issued by every member. If the ITD declares member j to be a threat, the logistics node immediately sets member j's GLAD-defined global and local database access privileges to *none* and then sends a message to every other member

stating that the ITD has declared member j to be a threat. Because all of these other members would then be aware of the threat, they would need to vote on whether member j should be separated from the confederation or not.

4. Identifying the Most Destructive Traffickers

The simulator consists of (a) submodels of the decision making of several groups, and (b) a submodel of the ecosystem affected by these groups. Group submodels in the simulator lack a spatial location input node and hence do not indicate in their decision option output where this option will be executed. For instance, a group submodel's decision to poach an animal does not carry with it a spatial location for that poaching action.

For a poaching decision in particular, the solution to this deficiency that is implemented here is as follows. For a particular group, estimate the spatial location of a decision to poach by finding the temporally closest observed region that experienced a previous poaching action by that group. This procedure for estimating the region where a poaching action occurred is also used to spatially locate group-generated poaching actions that are predicted by the model to occur in the future.

Before being used to support a wildlife trafficking investigation, the simulator is statistically fitted via *consistency analysis* (CA) (Haas, 2024a) to a *political-ecological* data set. This data set is composed of (a) an *actions history* data set collected via a STAR compliant protocol (Haas, 2024b), and (b) an ecological data set. Then, this fitted simulator is run forward from the present time to a *planning horizon date* and the extinction risk at that time point is computed for each region. Using the formula for extinction risk given in Haas and Ferreira (2016), a region's local extinction risk depends in-part on its poaching rate through time, habitat availability through time, and prey availability through time.

After entering this region-risk information into their FWCIDBMS, members issue queries to find traffickers associated with regions having high local extinction risks and who are also enjoying high social network influence as expressed by their *eigenvector centrality*. This social network analysis measure is computed within the WCIT's social network model module. See Haas and

Ferreira (2015) for a review of social network theory and associated measures as applied to the analysis of criminal intelligence.

The confederation then assigns the most influential of these traffickers to their Detain list. Next, based on social network computations, the confederation assigns to their Surveil list, *Rising Stars* (traffickers who are predicted to move into WTS leadership roles), and a *puppet master* (an influential trafficker attempting to hide their presence from law enforcement). Any pending trafficking actions detected by the confederation's intelligence-gathering are entered into their Interdict list. Finally, the confederation shares these three lists with law enforcement (governmental wildlife crime control agencies and international organizations pursuing prosecutions of wildlife traffickers).

In addition to the Detain, Surveil, and Interdict lists, this *actionable intelligence report* contains a *network resiliency* index for the WTS (a measure of how fast the syndicate's functionality can recover from a series of trafficker arrests).

4.1. Estimating the syndicate's Rising Stars and resiliency

The social network model module of the WCIT assumes that the confederation gathers evidence on the WTS at three different time points. Intelligence gathered at the first time point is used to find out the size, connectivity, and assets of the current, undisturbed WTS. Next, the confederation quietly watches the network for several weeks and at the end of that period, observes its size and connectivity again. Then, the confederation recommends to law enforcement those WTS traffickers to detain and surveil along with those near-future trafficking actions to interdict. Finally, some weeks after these arrests, the confederation gathers information on the size and connectivity of the recovering WTS. Call these three time points, t_1, t_2, and t_3, respectively.

Let $EC(p,t)$ be trafficker p's eigenvector centrality at time t. Trafficker p is a Rising Star if (a) $EC(p,t)$ is larger than the median eigenvector centrality of all traffickers in the WTS network at time t; and (b) $EC(p,t_2) > EC(p,t_1)$.

Let $CI(t)$ be a measure of a social network's *connectedness* at time t. Connectedness is one way to measure a social network's *functionality*. Let NRI be a measure of a social network's resiliency defined to be proportional to how quickly a social network recovers 90% of its functionality after removal of some of its traffickers.

One quantitative definition of $CI(t)$ is the largest eigenvalue of the social network's *link weight matrix*. And hence, one way to define *NRI* is to set it equal to $1/(t_3 - t_2)$ when $CI(t_3) = 0.9CI(t_2)$. This definition is operationalized by setting *NRI* to $CI(t_3)/((t_3 - t_2)0.9CI(t_2))$ when $CI(t_3) < 0.9CI(t_2)$ and declaring it to be at least $1/(t_3 - t_2)$, otherwise.

These two social network measures and the three-time-point strategy for attacking a WTS are all new.

4.2. The Detain list's arrest sequence computation

Arrest-priority is assigned to those traffickers who both reside in regions of predicted high local extinction risk and who have high eigenvector centrality. This prioritization is implemented by performing a two-level sort of all traffickers into a recommended sequence of arrests referred to as the *Bilevel Optimal Arrest Sequence*. Sort level 1 is a descending sort of all traffickers by the local extinction risk of the region of their residence. The second level sort is a descending sort on their eigenvector centrality at t_1. The Detain list is this arrest sequence. As detailed below, this arrest sequence is optimal in the sense that it is the solution to a *bilevel optimization problem*.

Because all traffickers in the confederation's database are included in this list, ecosystem damage is always the first priority when law enforcement arrests the first n traffickers from this list no matter what the value of n is. Depending on the political situation, law enforcement may have enough resources to arrest a large number of traffickers.

In summary, the Bilevel Optimal Arrest Sequence is produced by coupling a statistically fitted political-ecological model that hosts a preselected species to a social network model of the WTS that is harvesting that species. This coupling is new.

4.2.1. Mathematical form of the Bilevel Optimal Arrest Sequence

Blondel et al. (2020) show that ranking is a linear programming optimization problem. Let σ be a permutation of the integers 1 through n, and σ^{-1} be its *inverse permutation*. Letting j_i be the i^{th} entry in σ, the j^{th} entry in σ^{-1} is i. The problem's objective function is equation (4) in Blondel et al. (2020):

$$r(\boldsymbol{\theta}) = \underset{\mathbf{y} \in \mathcal{P}(\boldsymbol{\rho})}{\arg\max} \langle \mathbf{y}, -\boldsymbol{\theta} \rangle \tag{3}$$

where $\boldsymbol{\rho} = (n, n-1, \ldots, 1)$ is the *reversing permutation* vector, and $r(\boldsymbol{\theta}) = \sigma^{-1}(\boldsymbol{\theta})$, i.e., $r(\boldsymbol{\theta})$ is the vector of *ranks* that reorders the received scores, $(\theta_1, \ldots, \theta_n)$ into descending order: $\theta_{1_\sigma} > \ldots > \theta_{n_\sigma}$. The problem's two constraints are that $\boldsymbol{\theta} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathcal{P}(\boldsymbol{\rho})$ where $\mathcal{P}(\boldsymbol{\rho})$ is the *permutahedron* induced by $\boldsymbol{\rho}$. Here, a score is either extinction risk or eigenvector centrality.

Hence, the ordering of traffickers to arrest that is delivered by the Bilevel Optimal Arrest Sequence's two-level sort is the *bilevel optimal* (Jin and Yang, 2023) solution to a bilevel optimization problem where the first level sort on extinction risk is the "leader," and the second level sort on eigenvector centrality is the "follower." This Bilevel Optimal Arrest Sequence is new.

For example, say that traffickers 1, 2, and 3 all of whom reside in a high extinction risk region, possess eigenvector centrality scores: $\boldsymbol{\theta}' = (0.4, 0.2, 0.7)$, respectively. Then $r(\boldsymbol{\theta}) = (2, 3, 1)$. This means that trafficker 1 has rank 2, trafficker 2 has rank 3, and trafficker 3 has rank 1. Hence, the Bilevel Optimal Arrest Sequence is to arrest trafficker 3 first, then trafficker 1, and finally trafficker 2.

4.2.2. Discussion

Three points concerning this integration of political-ecological modelling and criminal intelligence need to be emphasized.

1. The fundamental challenge in biodiversity conservation is not the reduction of poaching but rather, the avoidance of local extinction events. This is why regions are prioritised by their local extinction risks rather than by their poaching rates.
2. These per-region extinction risks are generated by the simulator rather than by analysis of the confederation's associated social network model of those traffickers contained in the confederation's FWCIDBMS. And further, a political-ecological model is required in addition to a political-ecological data set in order to compute extinction risks at a future time point.
3. Analysts may not be trained in ecology or in wildlife management and hence, need a single, quantitative measure of ecological damage that they can incorporate into their criminal investigations. The ecologically sound, local extinction risk of a preselected species is one such measure.

5. Conserving the Cheetah

Many private, for-profit firms possess expertise in pursuing financial fraud investigations. Indeed, most insurance companies have an in-house *special investigation unit* (SIU) whose sole purpose is to investigate insurance fraud. Staff within such units include analysts experienced in conducting criminal investigations that include the detection of financial irregularities from online sources. Such a firm would be in a position to assign one of their existing fraud investigation units to wildlife trafficking investigations. Call this wildlife trafficking investigations effort, the firm's *biodiversity project*.

A firm could fund this project with revenue from a *biodiversity premium* that they would charge in addition to the regular price of one of their products or services. Haas (2022) calls such a product or service a *biodiversity offering*. The firm would market their biodiversity offering to customers who are concerned about biodiversity loss. The Intel Kit of Haas (2026) provides guidance and software to support a firm's efforts to develop such a business venture.

The following sections describe how a wildlife trafficking investigations project could help conserve the East African cheetah.

5.1. The biodiversity offering and biodiversity project

Say that a hypothetical insurance firm has chosen one of their auto insurance policies for their biodiversity offering. Using revenue from the offering's biodiversity premium, this firm decides to focus on combatting cheetah trafficking where these animals are most populous: East Africa. The cheetah is listed as Vulnerable on the IUCN Red list and as Endangered by the Namibian government (Milloway, 2025).

One form of such trafficking involves seizing live cheetah cubs in their den while their mother is away hunting. The few cubs who survive transport, are sold to private parties who desire an exotic pet (Tricorache et al., 2021). Countries where these seizures occur include Kenya, Tanzania, and Uganda (Tricorache and Stiles, 2021). Many of these transactions are arranged using social media platforms (Moneron and Nelwamondo, 2024). The East African cheetah population is also reduced by local farmers shooting adult cheetahs to protect their livestock.

The firm reaches this decision in-part because much of the trafficking in cheetah is international wherein shipments originating in East Africa have final delivery locations in the Middle East and in the United States. Such international trafficking gives the firm's SIU opportunities to gather intelligence from many sources on shipments and the criminals managing those shipments. These sources include social media sites and mobile phone networks.

The project consists of ongoing investigations of cheetah poaching events, cheetah poachers, cheetah cub shipments, cheetah body parts shipments, and the traffickers who (1) buy cheetahs and/or their body parts from poachers, (2) arrange transport of the ensuing shipments, and (3) arrange final retail sales of such shipments to consumers. Intelligence gathered in the course of these investigations is used to create a set of recommended law enforcement actions that is shared with the Kenya Wildlife Service (KWS), the Tanzania Wildlife Management Authority (TAWA), and other law enforcement agencies. These actions are conveyed in the Detain, Surveil, and Interdict lists as described above.

The project is implemented by leveraging current capabilities of the insurance firm's SIU. Specifically,

1. The firm assigns four SIU investigators part-time to the project. These investigators each bill one-third of their time to this project.
2. The firm joins a confederation of analysts. This confederation maintains an FWCIDBMS in order to allow members to share with each other, intelligence on traffickers and cheetah shipments. Further, this firm volunteers to maintain the logistics node of the confederation's FWCIDBMS.
3. The firm purchases a secure hardware/software package to run this logistics node.
4. Finally, the firm hires and deploys a four-person team to Nairobi, Kenya. This team consists of two analysts, an office manager, and an information technology specialist. This team gathers evidence that can only be acquired by intelligence-gathering methods that are deployed on the ground in East Africa. These two analysts feed such intelligence to the confederation's FWCIDBMS.

5.2. Monitoring program and cheetah abundance estimation

The confederation needs to statistically fit their cheetah-hosting political-ecological model to a combined actions history data set and an ecological data set. Here, this ecological data set consists of real-time sightings of East African cheetah. This sightings data is streamed to the confederation's FWCIDBMS.

Acquisition of sightings data in real-time requires the cooperation of East African conservation agencies. This cooperation is won through the services of the firm's *liaison consultant*. See Haas (2022) for a discussion of why this consultant is critical to the success of any in-country biodiversity project. And see Liaison (2025) for a step-by-step example of setting up a liaison office in a country that hosts a preselected species.

Once this sightings data is acquired, the work of Mallick (2023) can be followed as an example of using a *capture-recapture* statistical estimator to estimate the abundance of a terrestrial predator. Ferreira et al. (2020) take a continuous-time approach to this estimation challenge. A SAS code for such a computation with an accompanying example data set is available at Haas (2026).

5.3. Cheetah-hosting political-ecological system simulator

The agent/individual-based model of the cheetah-hosting political-ecological system consists of the following submodels:

1. Kenya pastoralists, and Tanzania pastoralists
2. Kenya rural residents, and Tanzania rural residents
3. KWS and TAWA
4. The presidential office of Kenya, and the presidential office of Tanzania
5. A conservation-focused *nongovernmental organization* (NGO) operating in both of these two countries
6. A spatio-temporal, individual-based submodel of cheetah abundance across Kenya and Tanzania.

See Haas (2025b) for the architecture, causal flow, and decision making mechanisim of the above group submodels. The cheetah abundance submodel is spatio-temporal because it computes an estimate of cheetah abundance for

each politically-defined region in Kenya and Tanzania at each week over a specified interval of years.

5.3.1. Extinction probabilities and extinction risk computations

Extinction probability is computed by finding the fraction of the number of realizations that give an abundance of less than thirteen individuals at the target date. A minimum viable population (MVP) size of 13 is in agreement with the range of 11 to 14 individuals given in Hilbers et al. (2017).

The formula for extinction risk at a future time point, t that is given in Haas and Ferreira (2016) is:

$$R(t) = L(t)P(\text{extinct at } t) \tag{4}$$

where the non-use loss due to extinction is $L(t) = (1 - 0.035)^t$ with 0.035 being the discount rate.

5.4. Data sets

5.4.1. Actions history

A total of 1272 actions from 2009 through 2025 have been collected via the STAR compliant protocol developed by Haas (2024b). This actions history data set is summarized in Table 3. In particular, this data contains cheetah poaching actions that have occurred within specific regions.

Stories file name	Year(s)	Number of stories
ef9-13.txt	2009-2013	4178
ef13-14.txt	2013-2014	1655
ef14-15.txt	2014-2015	2443
ef15-16.txt	2015-2016	9293
ef16-19.txt	2016-2019	10910
ef19.txt	2019	9806
ef19-21.txt	2019-2021	8542
ef20211231.txt	2021	11623
ef20221231.txt	2022	30017
ef20230221.txt	2023	3016
ef20230529.txt	2023	2821
ef20231008.txt	2023	4605
ef20240111.txt	2024	2881
ef20240327.txt	2024	3173
ef20240609.txt	2024	3044
ef20240827.txt	2024	3508
ef20241112.txt	2024	3466
ef20250212.txt	2025	2391
ef20250315.txt	2025	1393
ef20250516.txt	2025	380
ef20250715.txt	2025	2405
ef20250906.txt	2025	2658

Table 3: Summary of the actions history data set. Stories files cover the years 2009 through 2025. Action detection is performed with the `parse_stories` relation of the FWCIDBMS.

5.4.2. Ecological data

Cheetah sightings data is typically collected by field ecologists running camera traps or observing cheetah spoor. Based on these data-collection methods, the World Population Review (2025) reports 938 cheetah in Tanzania and 715 in Kenya (for a total of 1653). A more detailed, regional cheetah abundance data set for regions in Kenya and Tanzania is based on observations reported in the official report on the status of cheetah written by the International Union for the Conservation of Nature (IUCN) (IUCN/SSC Cat Specialist Group, 2022). This data set along with associated patch adjacency information is contained in the file, `cheetahpatches.dat` (Table 4).

Patch ID	Region	Cheetah abundance	Number of adjacent patches	Adjacent patches
		Kenya		
1	Turkana	33	1	2
2	Mandera-Marsabit	175	1	1
3	Tsavo	650	1	6
		Tanzania		
4	Serengeti	600	1	6
5	Katavi-Ugalla	55	1	7
6	Maasai-steppe	47	1	4
7	Ruaha	184	1	5

Table 4: The regional cheetah abundance input file, `cheetahpatches.dat`. These 2022 values are used for estimating abundance in the year 2025. The file includes inter-region adjacency information.

This file's inter-region adjacency relationships are read from publicly accessible maps (Wikipedia, 2025a; Masai Mara Travel, 2025; Wikipedia, 2025b). Adjacency information is necessary to model cheetah movements across region boundaries.

5.4.3. Hypothetical criminal intelligence gathered on the WTS

A hypothetical data set on a WTS is created for the purposes of illustrating how a confederation of analysts and their FWCIDBMS would help reduce trafficking in cheetahs and their body parts. This hypothetical data set (file `cheetah_wts.dat`) (Figure 2) contains the connectivity between traffickers in an East African WTS that trades in live cheetahs and cheetah body parts.

Intelligence on a real-world WTS would, of course, be preferable. It has been this author's experience, however, that acquiring such a data set for research purposes often requires the researcher gain the trust of a wildlife crime investigation unit before such a unit gives access to their confidential criminal intelligence data. Before being handed to a researcher, such data needs to be *anonymized*, *de-identified*, or *obfuscated* in order remove trafficker identities (Krishnamoorthy, 2025). This is done to protect the investigation unit and the researcher from trafficker reprisals. Such an embedding by this author was indeed successful during the investigation of a WTS engaged in

the trafficking of rhino horn from a population of white rhinoceros *Ceratotherium simum* (Haas and Ferreira, 2015). This author is not currently interacting with a wildlife crime investigation unit.

```
linksfiletype nmtimepts
5             3
time_point nmplayers nmlevels nmlinks
0.0        10        4        11
name level town region country nmvehicles vehicles
r1 4 town1 tsavo        country1 0
m4 2 town4 turkana       country3 0
m3 2 town1 maasai_steppe   country1 0
r2 4 town1 maasai_steppe   country1 0
t2 3 town3 serengeti         country2 0
t1 3 town1 tsavo          country1 1 lu7
h22 1 town1 serengeti       country1 0
m1 2 town1 turkana      country1 0
m2 2 town1 turkana      country1 0
h12 1 town1 turkana     country1 0
player 1 player 2 type
h12 m2 call
m2 h22 call
m2 m1 shipment
h22 m1 shipment
t1 m1 transfer
t2 t1 call
t2 r2 call
t2 r1 call
t2 m4 call
t2 m3 call
m3 m3 call
time_point nmplayers nmlevels nmlinks
17.0       10        4        13
name level town region country nmvehicles vehicles
r1 4 town1 tsavo        country1 0
m4 2 town4 turkana       country3 0
m3 2 town1 maasai_steppe  country1 0
r2 4 town1 maasai_steppe  country1 0
t2 3 town3 serengeti        country2 0
t1 3 town1 tsavo         country1 1 lu7
h22 1 town1 serengeti       country1 0
m1 2 town1 turkana      country1 0
m2 2 town1 turkana      country1 0
h12 1 town1 turkana     country1 0
player 1 player 2 type
h12 m2 call
m2 h22 call
m2 m1 shipment
h22 m1 shipment
t1 m1 transfer
t2 t1 call
t2 r2 call
t2 r1 call
t2 m3 call
t2 m4 call
m3 m4 call
m2 m3 call
m2 m4 call
```

```
time_point nmplayers nmlevels nmlinks
24.7       9         4         9
name level town region country nmvehicles vehicles
r1 4 town1 tsavo        country1 0
m3 2 town1 maasai_steppe  country3 0
r2 4 town1 maasai_steppe  country1 0
t2 3 town3 serengeti        country2 0
t1 3 town1 tsavo         country1 1 lu7
h22 1 town1 serengeti       country1 0
m1 2 town1 turkana      country1 0
m2 2 town1 turkana      country1 0
h12 1 town1 turkana     country1 0
player 1 player 2 type
m2 h22 call
m2 m1 shipment
h22 m1 shipment
t1 m1 transfer
t2 t1 call
t2 r2 call
t2 r1 call
t2 m3 call
m2 m3 call
```

Figure 2: Criminal intelligence gathered by the confederation on the WTS operating in East Africa.

5.5. Results

5.5.1. Parameter estimation and local extinction risk predictions

The above actions history data set and the above ecological data set form a political-ecological data set. This data set is used to statistically estimate the parameters of the Kenya rural residents, and the Tanzania rural residents submodels via the CA statistical estimator of Haas (2024a). Due to computing resource limitations, only data from 2021 to 2025 is used to fit the model. CA increased the value of its statistical goodness-of-fit measure by 23.3% (Haas, 2025a). The fraction of observed actions matched by the fitted model is 0.509.

Next, this fitted model is run forward in time to the planning horizon year of 2030 in order to predict local cheetah extinction risks by region (Table 5). Only regions having an extinction probability of less than 1.0 are included in this Table. The reasoning for this is that if extinction is certain in a particular region, pursuing poachers there would not help to sustain the overall cheetah population.

Region	Extinction Probability	Extinction Risk
Kenya		
Turkana	0.275	0.229
Maasai-steppe	0.025	0.020
Tanzania		
Serengeti	0.025	0.020
Katavi-Ugalla	0.000	0.000
Mandera-Marsabit	0.000	0.000
Ruaha	0.000	0.000

Table 5: Predicted local cheetah extinction risks by region for the year 2030. Regions of certain extinction are excluded.

These region-risk results are entered into the confederation's social network model module in order to produce the actionable intelligence report that the confederation will share with law enforcement. This analysis is described next.

5.5.2. Construction of the actionable intelligence report

Using a social network model of the traffickers contained in their FW-CIDBMS, the confederation constructs their actionable intelligence report as shown in Figure 3. This report is generated by running the **id** *relations file*, `kentan.id` with the command

```
idalone kentan.id
```

at a Windows or Linux command prompt depending on where the WCIT has been installed.

```
----- Detain List (Based on Network at Time point 1) -----

 Simulator-SNA-generated Bilevel Optimal Arrest Sequence:
 Arrest_Priority  Extinction_Risk Eigenvector_Centrality  Player_Name
   1       00.229        00.383       m4
   2       00.229        00.199       m1
   3       00.229        00.199       m2
   4       00.229        00.122       h12
   5       00.020        00.298       t2
   6       00.020        00.199       h22

---- Surveil List (Based on Network at Time point 2) ------
SNA-generated Successor Prediction(s):
   r2 will succeed m4.

SNA-predicted Influential Player Attempting to Hide
(highest ratio of betweenness centrality-to-degree centrality):
     t1

SNA-predicted Rising Stars (Based on Time points 1 and 2):
   r1 is a Rising Star
   m4 is a Rising Star
   m3 is a Rising Star
   r2 is a Rising Star
   t2 is a Rising Star
   t1 is a Rising Star

------------------- Interdict List --------------------
1. Recommendation: Seize a boat that will be sailing along the
Somaliland coast in late September 2025.  This boat will be carrying
cheetah cubs for the exotic pet trade.

--- Network Resiliency Index (Recovery Time) ---
  (assumes arrests were made just after time point 2)
  Connectivity Index at latest time =    2.518
   Connectivity Index prior to arrests =    3.166
   Network Resiliency Index =   00.114 or about    8.713 weeks.
```

Figure 3: Final section of the actionable intelligence report built from an integration of simulator-computed local cheetah extinction risks and a social network analysis of the trafficker intelligence contained in the FWCIDBMS. This report also contains several social network analysis measures that support the report's Detain, Surveil, and Interdict lists. Identity and location information of the traffickers referred to in these lists is shared with law enforcement.

Figures 4 and 5 show the effects on the WTS due to the arrest recommended in the Detain list. These Figures indicate that the removal of trafficker `m4` (a middleman) reduces the network's connectivity and isolates trafficker `h12`. The WTS is expected to recover from this damage in about 8.713 weeks.

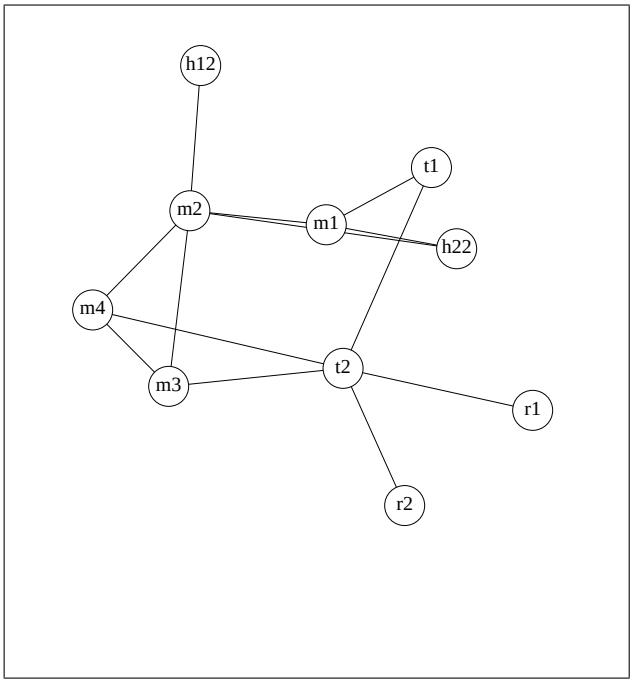

Figure 4: The syndicate's social network just before the arrest given in the Detain List is made.
Alt text: The Figure shows a social network. Player m2 is connected to h12, m4, m3, m1, and h22. Player m1 is connected to t1 and h22. Player m4 is connected to t2 and m3. Player t2 is connected to t1, r1, and r2.

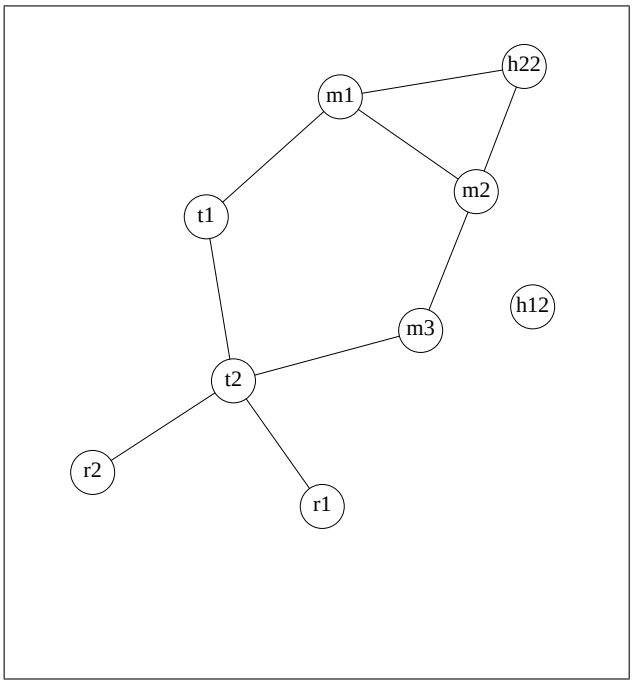

Figure 5: The syndicate's social network some weeks after the arrest of trafficker `m4`.
Alt text: The Figure shows a social network. Player t2 is connected to r2, r1, and m3. Player m1 is connected to h22 and m2. Player h12 is not connected to anyone.

An example of a trafficking action that would need interdiction is as follows. Say that the confederation hears from one of their informants about a shipment of cheetah cubs to be made by boat along the Somaliland coast. By including this planned shipment as an item in their Interdict list, the confederation would recommend that this shipment be seized. Procedurally, the confederation would send all three lists to many law enforcement authorities including the Somali Coast Guard. Then, the Somali Coast Guard would physically execute the recommended interdiction.

Such an incident actually happened: Eleven cheetah cubs headed for the exotic pet trade were seized by the Somali Coast Guard in late September 2025 (Associated Press, 2025). These cubs were being transported by boat along the Somaliland coast when they were intercepted by the Somali Coast Guard.

6. Discussion

There is reason to believe that the most effective way to curb the illegal trade in a particular species is to focus on interdicting those traffickers who reside in its home range and who are directly involved in its poaching and transport (Felbab-Brown, 2018). If true, this strategy for fighting wildlife crime is congruent with one of the main points of this article: Anti-trafficking efforts should concentrate on species-hosting regions that are predicted to have high local extinction risks in the future.

6.1. Shortcomings

1. As discussed by Haas (2023), efforts to curb wildlife trafficking need to increase at least ten-fold in order to stem the rapid loss of global biodiversity (circa 2026). Intelligence-sharing agreements such as the confederation-approach developed in this article and in Haas (2023) are needed to help break the international syndicates that fuel this destruction of wildlife. But there are severe trust and financial roadblocks to overcome before such sharing can occur.

   The ITD is a start towards convincing the intelligence community that the sharing of wildlife trafficking intelligence need not compromise their security – but much more needs to be done before such analysts would actually be willing to share their hard-won, highly confidential criminal intelligence.
2. The biodiversity offering approach of Haas (2022) is also a start towards the funding of international wildlife trafficking investigations. But the needed increase in funding for increased wildlife crime investigations is so great that it is difficult to see how such funding increases could be accomplished within present funding mechanisms. Currently, such mechanisms are mostly based on private donations and the relatively low levels of taxpayer-support given to wildlife crime control agencies.
3. Within-database methods of detecting insider threats such as the ITD, are not useful at detecting those users who may become a threat in the near-future. To do this, it would be necessary to monitor each member's personal finances, contacts, and ideally, evidence of the member

accepting bribes. Doing so would help prevent insider attacks from happening in the first place.

6.2. Future work

An alternate categorical encoding method is planned for future development – that of *complex encoding* (Kunanbayev et al., 2021). This method may provide superior anomaly detection performance for only a two-fold increase in its memory requirement over ordinal encoding. The reasons for this choice are as follows.

First, Bolikulov et al. (2024) do not find critical differences in accuracy of several categorical encoders including ordinal encoding, one-hot encoding (Hancock and Khoshgoftaar, 2020), and hashing encoding. Second, the application of ordinal encoding herein is solely for the purposes of detecting a change in a collection of attribute values from one set of query results to those in a subsequent set of query results. Because here, detecting such a change is the only requirement of a categorical encoder algorithm, the only relevant test of a good encoder is whether it can support such detections. Veracity of the original categories in the encoded space is not relevant to this application. Rather, here, an encoder needs to be judged on whether it can detect a difference in the encoded space over many possible new collections of attribute-category values contained in a new query result. For the case of ordinal encoding, the most difficult collection to detect a change in is when the new collection of encoded category values are so embedded between the ordinal values of the old query result that neither the ordinal median, ordinal IQR, nor the size of the new query result are different from those of the old query result. When viewed as a sampling experiment, having all three of these query result summary measures be the same between the old and the new is a joint event of these three random variables taking on the same values between the old result and the new result. The probability of a joint event happening that is composed of a number of events, becomes small as the number of these events increases. This result holds even when the random variables in question are dependent. Adding more summary measures increases the number of events that need to happen jointly before the encoder fails to show a difference between the content of an old query result and the

content of a new query result. To this end, the use of complex encoding will double the number of summary measures used to detect differences.

The proof of this decreasing probability is straightforward. Let $A_1, \ldots, A_n$ be events. Let $I_n = \bigcap_{i=1}^{n} A_i$ and $I_{n+1} = \bigcap_{i=1}^{n+1} A_i$. Then, by the monotonicity property of probability, $P(I_{n+1}) \leq P(I_n)$ (because I_{n+1} is the same or smaller than I_n). Hence, $P(\bigcap_{i=1}^{n+1} A_i) \leq P(\bigcap_{i=1}^{n} A_i)$.

7. Conclusions

This article has described a freely available WCIT that can help a confederation of analysts curb wildlife trafficking. This WCIT consists of three modules: An FWCIDBMS, a political-ecological model of the system that hosts a preselected species, and a social network model of the WTS that is harvesting this preselected species. The latter two modules enable a confederation to focus their investigations onto those traffickers most responsible for the highest local extinction risks of this preselected species. This focus is enabled by integrating a model of a political-ecological system that hosts the preselected species with a social network model of the attacking WTS. Operationalizing this focus on high-extinction-risk traffickers within a wildlife trafficking investigation is new.

The most extensive actions history data set to-date (circa 2026) on cheetah trafficking has been collected by the author using a STAR compliant protocol (Haas, 2024b). This data set has been used herein to show how this integration guides an investigation onto those traffickers most responsible for driving a preselected species towards extinction.

This article has also presented and tested a new, data-centric algorithm that protects the FWCIDBMS module from insider attacks. This threat detection system is critical for convincing a diverse, multinational group of analysts to voluntarily join a confederation that gathers, shares, and analyzes criminal intelligence to help curb wildlife trafficking.

To slow the rapid loss of biodiversity across the globe, it is crucial for wildlife trafficking investigations to be guided by credible models of local species extinction risks. One way to provide such guidance has been presented in this article. Such political-ecological guidance of large-scale wildlife

trafficking investigations may be the only way for such investigations to save those species who are on the brink of extinction.

8. Conflicts of Interest

The author declares that he has no competing interests.

9. Funding

This work received no funding support.

10. Data Availability

All source code and input files needed to run the WCIT on the examples described herein are contained in the three files `java_source.zip`, `scripts.zip`, and `data.zip`. The first of these files contains the toolkit's JAVA source code – namely, the program **id**. The second file contains all needed Linux® shell scripts and Windows® batch files. The third file contains model definition files and all of the political-ecological data analyzed in this article. These three files are freely available at either `www.profitablebiodiversity.com/software` or the Zenodo repository (see `https://about.zenodo.org/policies/`). Access Zenodo versions with the following digital object identifiers (DOIs):

10.5281/zenodo.19653266 (April 19, 2026 version) or
10.5281/zenodo.19653265 (all versions resolving to the latest).

11. Author Contributions Statement

T.C.H. developed the concept, wrote all code used in the article, collected all of the data, ran all of the computations, wrote the article's text, and reviewed the article.

12. Acknowledgments

The author thanks the anonymous reviewers for their valuable suggestions.

References

Advance Datasec (2026), *Types of Insider Threats in Cyber Security and How to Detect Them,* https://advance-datasec.com/insider-threats-in-cyber-security/

Al-Shehari, T., Rosaci, D., Al-Razgan, M., Alfakih, T., Kadrie, M., Afzal, H., and Nawaz, R. (2024), "Enhancing Insider Threat Detection in Imbalanced Cybersecurity Settings Using the Density-Based Local Outlier Factor Algorithm," *IEEE Access*, 12: 34820-34834. doi: 10.1109/ACCESS.2024.3373694.

Anand, R., Mehrotra, K., Mohan, C. K., and Ranka, S. (1995), "Efficient Classification for Multiclass Problems Using Modular Neural Networks," *IEEE Transactions on Neural Networks,* 6(1): 117-124, January. DOI: 10.1109/72.363444.

Arif, S. A. B. and Wani, S. (2025), "The Theoretical Foundations and Literature Analysis a Hybrid Detection Technique Against Malicious SQL Attacks on Web Applications," *Journal of Information Systems Engineering and Management,* 10(35s): 1093-1100, e-ISSN: 2468-4376. https://www.jisem-journal.com/

Associated Press (2025), "Cheetah Cubs Destined for Illegal Trade in Exotic Pets Rescued in Somaliland," *Associated Press*, October 2. https://apnews.com/article/somaliland-cheetahs-rescued-gulf-exotic-pets-56c4ab013c1230770f3eb18782dc7fc7

Bao, H. and Gao, J. (2025), "Network Intrusion Detection Based on Improved KNN Algorithm," *Scientific Reports,* 15, 29842. https://doi.org/10.1038/s41598-025-14199-2.

Blondel, M., Teboul, O., Berthet, Q., and Djolonga, J. (2020), "Fast Differentiable Sorting and Ranking," (in) *Proceedings of the 37th International Conference on Machine Learning (ICML'20),* 119. JMLR.org, Article 89, pp. 950–959.

Bolikulov, F., Nasimov, R., Rashidov, A., Akhmedov, F., and Cho, Y.-I. (2024), "Effective Methods of Categorical Data Encoding for Artificial In-

telligence Algorithms," *Mathematics,* 12(2553), https://doi.org/10.3390/math12162553.

Castano, S., De Capitani di Vimercati, S., and Fugini, M. G. (1997), "Automated Derivation of Global Authorizations for Database Federations," *Journal of Computer Security,* 5(4): 271-301, DOI: 10.3233/JCS-1997-5402.

Chaurasia, A. K. (2023), "Tower Dumps: Tracking Wildlife Criminals," *WildHub*, Oct 29. https://wildhub.community/posts/tower-dumps-tracking-wildlife-criminals

Demeau, E., Vargas, M., and Jeffrey, K. (2019), "Wildlife Trafficking on the Internet: A Virtual Market Similar to Drug Trafficking?" *Revista Criminalidad,* 61(2): 101-112.

ECO-SOLVE (2024), "Monitoring online illegal wildlife trade," *Global Initiative Against Organized Transnational Crime,* https://globalinitiative.net/analysis/monitoring-online-illegal-wildlife-trade/

Felbab-Brown, V. (2018), *To Counter Wildlife Trafficking, Local Enforcement, Not En-Route Interdiction, is Key, Brookings,* January 19. https://www.brookings.edu/articles/to-counter-wildlife-trafficking-local-enforcement-not-en-route-interdiction-is-key/

Ferreira, S. M., Beukes, B. O., Haas, T. C., and Radloff, F. G. T. (2020), "Lion (Panthera leo) Demographics in the Southwestern Kgalagadi Transfrontier Park," *African Journal of Ecology*, 58(2): 348-360. DOI: 10.1111/aje.12728.

Goyal, A., Han, X., Wang, G., and Bates, A. (2023), "Sometimes, You Aren't What You Do: Mimicry Attacks against Provenance Graph Host Intrusion Detection Systems," *Network and Distributed System Security (NDSS) Symposium 2023,* 27 February - 3 March 2023, San Diego, CA, USA https://dx.doi.org/10.14722/ndss.2023.24207

Gramer R. (2017), "Israel Changed Intelligence Sharing with U.S. After Trump Comments to Russians," *Foreign Policy*, 24, May.

https://foreignpolicy.com/2017/05/24/israel-changed-intelligence-sharing-with-u-s-after-trump-comments-to-russians/

Haas, T. C. (2026) *Profitable biodiversity website.* https://profitablebiodiversity.com.

Haas, T. C. (2025a), "Using Political-Ecological Models to Sustain Biodiversity," under review at *Scientific Reports*. Preprint available at www.profitablebiodiversity.com.

Haas, T. C. (2025b), "A Technology-Based Business Plan for Profitably Curbing Wildlife Trafficking," *Sustainability Technology* (SusTech 2025), Santa Ana, California, April 20-23. See page 50 of: https://ieee-sustech.org/wp-content/uploads/sites/261/2025/04/SusTech-2025-Program-Guide.pdf

Haas, T. C. (2024a), "Models Vetted Against Prediction Error and Parameter Sensitivity Standards Can Credibly Evaluate Ecosystem Management Options," *Ecological Modelling*, 498, December, 11090 ("decreases" should be "increases" in the Graphical Abstract). DOI: 10.1016/j.ecolmodel.2024.110900.

Haas, T. C. (2024b). Protocol to Discover Machine-Readable Entities of the Ecosystem Management Actions Taxonomy. *STAR Protocols*, Cell Press, Elsevier, 5(2), 103125: 1-12. DOI: 10.1016/j.xpro.2024.103125.

Haas, T. C. (2023), "Adapting Cybersecurity Practice to Reduce Wildlife Cybercrime," *Journal of Cybersecurity*, 9(1): 1-20. DOI: 10.1093/cybsec/tyad004.

Haas, T. C. (2022), "Profitable Biodiversity," *Cogent Social Sciences*, 8(1): 1-24. DOI: 10.1080/23311886.2022.2116814.

Haas, T. C. (2021), "The First Political-Ecological Database and its Use in Episode Analysis," *Frontiers in Conservation Science, section: Planning and Decision-Making in Human-Wildlife Conflict and Coexistence*, 2:707088. DOI: 10.3389/fcosc.2021.707088.

Haas, T. (2020), "Developing Political-Ecological Theory: The Need for Many-Task Computing," *PLOS ONE*, November 24. DOI: 10.1371/journal.pone.0226861.

Haas, T. C. and Ferreira, S. M. (2016), "Conservation Risks: When Will Rhinos be Extinct?" *IEEE Transactions on Cybernetics*, 46(8): 1721-1734. Special issue on Risk Analysis in Big Data Era. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7236914.

Haas, T. C. and Ferreira, S. M. (2015), "Federated Databases and Actionable Intelligence: Using Social Network Analysis to Disrupt Transnational Wildlife Trafficking Criminal Networks," *Security Informatics,* 4:1. DOI: 10.1186/s13388-015-0018-8. http://www.security-informatics.com/content/4/1/2 http://www.springer.com/-/4/0d7808225b2a4876986ead314e72ee99

Hancock, J. T. and Khoshgoftaar, T. M. (2020), "Survey on Categorical Data for Neural Networks," *Journal of Big Data,* 7(28). https://doi.org/10.1186/s40537-020-00305-w.

Hand, D. J., Christen, P., and Kirielle, N. (2021), "F*: an Interpretable Transformation of the F-measure. *Machine Learning,* 110: 451–456. https://doi.org/10.1007/s10994-021-05964-1.

Hilbers, J. P., Santini, L., Visconti, P., Schipper, A. M., Pinto, C., Rondinini, C. and Huijbregts, M. A. J. (2017), "Setting Population Targets for Mammals Using Body Mass as a Predictor of Population Persistence," *Conservation Biology,* 31: 385-393. https://doi.org/10.1111/cobi.12846.

Inayat, U., Farzan, M., Mahmood, S., Zia, M. F., Hussain, S., and Pallonetto, F. (2024), "Insider Threat Mitigation: Systematic Literature Review," *Ain Shams Engineering Journal,* 15(12): 103068. https://doi.org/10.1016/j.asej.2024.103068.

IUCN/SSC Cat Specialist Group (2022), *Cheetah Acinonyx jubatus* https://www.catsg.org/living-species-cheetah.

Jin, H., and Yang, X. (2023), "Bilevel Optimal Sizing and Operation Method of Fuel Cell/Battery Hybrid All-Electric Shipboard Microgrid," *Mathematics,* 11(12): 2728, https://doi.org/10.3390/math11122728

Koehler, G., Schmidt-Küntzel, A., Marker, L., and Hobson, K. A. (2023), "Delineating Origins of Cheetah Cubs in the Illegal Wildlife Trade: Improvements Based on the Use of Hair $\delta^{18}O$ Measurements," *Frontiers in Ecology and Evolution,* 11. DOI: 10.3389/fevo.2023.1058985.

Kosaraju, N., Sankepally, S. R., Mallikharjuna Rao, K. (2023). "Categorical Data: Need, Encoding, Selection of Encoding Method and Its Emergence in Machine Learning Models - A Practical Review Study on Heart Disease Prediction Dataset Using Pearson Correlation," (in) Saraswat, M., Chowdhury, C., Kumar Mandal, C., Gandomi, A.H. (eds.) *Proceedings of International Conference on Data Science and Applications,* Lecture Notes in Networks and Systems,551. Springer, Singapore. https://doi.org/10.1007/978-981-19-6631-6_26

Krishnamoorthy, V. M. (2025), "Data Obfuscation Through Latent Space Projection for Privacy-Preserving AI Governance: Case Studies in Medical Diagnosis and Finance Fraud Detection," *JMIRx Med.*, March 12, 6: e70100. doi: 10.2196/70100.

Kul, G., Upadhyaya, S., and Hughes, A. (2020), "An Analysis of Complexity of Insider Attacks to Databases," *ACM Transactions on Management Information Systems,* 12(1): Article 4 (December). DOI: 10.1145/3391231.

Kunanbayev, K., Temirbek, I., and Zollanvari, A. (2021), "Complex Encoding," *International Joint Conference on Neural Networks (IJCNN),* Shenzhen, China, pp. 1-6, doi: 10.1109/IJCNN52387.2021.9534094.

Liaison (2025), *When To Set Up A Liaison Office in India 2024.* https://www.maiervidorno.com/blog/when-to-set-up-a-liaison-office-in-india/

Li, J., Wang, G. A., and Chen, H. (2011), "Identity Matching Using Personal and Social Identity Features," *Information Systems Frontiers,* 13(1): 101-113, March. https://doi.org/10.1007/s10796-010-9270-0

Lindauer, B. (2020). *Insider Threat Test Dataset,* Carnegie Mellon University. Dataset. https://doi.org/10.1184/R1/12841247.v1.

Mallick, J. K. (2023), "Conservation Status of Bengal Tiger Panthera tigris tigris in the Earth's Only Mangrove Tigerland: A Review of Efforts and Challenges," *Probe - Animal Science,* 5(1): 1-27. DOI: 10.18686/pas.v5i1.1777.
https://probe.usp-pl.com/index.php/PAS/article/viewFile/1777/1688

Márquez, M. C. (2025), "Wildlife Trafficking Goes Digital and Conservationists Are Racing To Catch Up," *Forbes*, August 19. https://www.forbes.com/sites/melissacristinamarquez/2025/08/19/wildlife-trafficking-goes-digital-and-conservationists-are-racing-to-catch-up/

Marquis, Y. A. (2024), "From Theory to Practice: Implementing Effective Role-Based Access Control Strategies to Mitigate Insider Risks in Diverse Organizational Contexts," *Journal of Engineering Research and Reports*, 26(5): 138–154.
DOI: 10.9734/jerr/2024/v26i51141.

Masai Mara Travel (2025), *Map of Kenya.*
https://www.masaimara.travel/map-of-kenya.php

Mathew, S., Petropoulos, M., Ngo, H. Q., and Upadhyaya, S. (2010), "A Data-Centric Approach to Insider Attack Detection in Database Systems," In: Jha, S., Sommer, R., Kreibich, C. (eds.) *Recent Advances in Intrusion Detection (RAID 2010). Lecture Notes in Computer Science,* 6307. Springer, Berlin, Heidelberg.
DOI: 10.1007/978-3-642-15512-3_20.

Milloway, O. (2025), *TWS 2024: Lead is 'Silently Poisoning' Captive Cheetahs, The Wildlife Society,* June 30. https://wildlife.org/tws-2024-lead-is-silently-poisoning-captive-cheetahs/

Moneron, S. and Nelwamondo, C. (2024), *Social Media Stimulating Trade in Cheetahs as Pets, Say New Data, Traffic,* March. https://www.traffic.org/publications/reports/online-live-cheetahs-trade-2024/.

Muñiz, D. Á, Miguel, L. P., Miguel, Muñoz, A. M., Larriva-Novo, X., Alvarez-Campana, M., and Rivera, D. (2026), "Design and Generation of a Dataset for Training Insider Threat Prevention and Detection Models: The SPEDIA dataset," *Computers and Security,* 161: 104743. https://doi.org/10.1016/j.cose.2025.104743.

Paul, A., Sharma, V., and Olukoya, O. (2024), "SQL Injection Attack: Detection, Prioritization and Prevention," *Journal of Information Security and Applications,* 85: 103871. https://doi.org/10.1016/j.jisa.2024.103871.

Raywood, D. (2018), "Top Ten Cases of Insider Threat," *Infosecurity Magazine,* 25 December.
https://www.infosecurity-magazine.com/magazine-features/top-ten-insider-threat/

Redivo, E., Viroli, C., and Farcomeni, A. (2023), "Quantile-Distribution Functions and Their Use for Classification, with Application to Naïve Bayes Classifiers," *Statistics and Computing,* 33(55). DOI: 10.1007/s11222-023-10224-4.

Roy, R., and Kumar, V. (2024), "An Analysis of Illegal Wildlife Trade with the Aid of Social Media and Prevention Strategies," *Journal of Wildlife and Biodiversity,* 8(1): 386-401. DOI: https://doi.org/10.5281/zenodo.10207005

Sardari, P., Badelu, N., Rajabipour, P., Mohammadi, A., Roberts, D. L., Kyle, G., and Farhadinia, M. S. (2026), "Characterizing the Illegal Trade of Carnivores on a Social Media Platform in Iran," *Biological Conservation,* 313(111521). DOI: https://doi.org/10.1016/j.biocon.2025.111521

SAS (2025), *PCTLDEF=3* (in) *Computing Quantiles, SAS/STAT 15.3 User's Guide.*
https://documentation.sas.com/doc/en/statug/15.3/statug_stdize_details03.htm

Setiono, R. and Hui L. K. (1995), "Use of a Quasi-Newton Method in a Feedforward Neural Network Construction Algorithm," *IEEE Transactions on Neural Networks,* 6(1): 273-277. DOI: 10.1109/72.363426.

Sharma, K., Barbosa, J. S., Roberts, S., Gondhali, U., Petrossian, G., Jacquet, J., Freire, J., and Chakraborty, S. (2025), "Descriptive Analysis of Online Wildlife Products Using Vision Language Models," In *Proceedings of the 2025 ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS '25)*, Association for Computing Machinery, New York, NY, USA, 461–472. DOI: https://doi.org/10.1145/3715335.3735484

Stringham, O. C., Maher, J., Lassaline, C. R., Wood, L., Moncayo, S., Toomes, A., Heinrich, S., Watters, F., Drake, C., Chekunov, S., Hill, K. G. W., Decary-Hetu, D., Mitchell, L., Ross, J. V., and Cassey, P. (2023), "The Dark Web Trades Wildlife, But Mostly for Use as Drugs," *People and Nature,* 5: 999–1009. DOI: https://doi.org/10.1002/pan3.10469

Tricorache, P. and Stiles, D. (2021), *Black Market Brief: Live Cheetahs, Global Initiative against Transnational Organized Crime*, September. https://globalinitiative.net/wp-content/uploads/2021/09/GITOC-ESAObs-Live-Cheetahs-Black-Market-Brief.pdf.

Tricorache, P., Yashphe, S., and Marker, L. (2021), "Global Dataset for Seized and Non-Intercepted Illegal Cheetah Trade (Acinonyx jubatus) 2010–2019, *Data in Brief,* 35(106848). DOI: 10.1016/j.dib.2021.106848. https://www.sciencedirect.com/science/article/pii/S2352340921001323.

Wang, R., Li, C., Zhang, K., and Tu, B. (2025), "Zero-Trust Based Dynamic Access Control for Cloud Computing," *Cybersecurity*, 8(12). DOI: 10.1186/s42400-024-00320-x.

Wikipedia (2025a), *Kenya Counties.* https://en.wikipedia.org/wiki/Counties_of_Kenya

Wikipedia (2025b), *Regions of Tanzania*.
https://en.wikipedia.org/wiki/Regions_of_Tanzania

World Population Review (2025), *Category: Environment*.
https://worldpopulationreview.com/country-rankings/cheetah-population-by-country

Wyatt, T., Miralles, O., Massé, F., Lima, R., Vargas da Costa, T., and Giovanini, D. (2022), "Wildlife Trafficking via Social Media in Brazil," *Biological Conservation,* 265, 109420, DOI: https://doi.org/10.1016/j.biocon.2021.109420