# The First Political-Ecological Database and Its Use in Episode Analysis

Timothy C. Haas*

Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Biodiversity loss is a consequence of socio-ecological processes. Observations on anthropogenic actions toward ecosystems coupled to observations on ecosystem metrics are needed to help understand these processes so that ecosystem management policies can be derived and implemented to curb such destruction. Such data needs to be maintained in searchable data portals. To this end, this article delivers a first-of-its-kind relational database of observations on coupled anthropogenic and ecosystem actions. This Ecosystem Management Actions Taxonomy (EMAT) database is founded on a taxonomy designed to support models of political-ecological processes. *Structured query language* scripts for building and querying these databases are described. The use of *episodes* in the construction of political-ecological theory is also introduced. These are frequently occurring sequences of political-ecological actions. Those episodes that test positive for causality can aid in improving a political-ecological theory by driving modifications to an attendant computational model so that it generates them. Two relational databases of political-ecological actions are described that are built from online news articles and published data on species abundance. The first concerns the management of the East African cheetah (*Acinonyx jubatus*) population, and the second is focused on the management of rhinoceroses (*Ceratotherium simum*) in South Africa. The cheetah database is used to study the political drivers of cheetah habitat loss, and the rhino database is used study the political drivers of rhino poaching. An EMAT database is a fundamental breakthrough because is provides a language for conservation science to identify the objects and phenomena that it is about. Therefore, maintaining political-ecological data in EMAT databases will advance conservation science and consequently, improve management policies that are based on that science.

Keywords: biodiversity loss, cheetah (*Acinonyx jubatus*), rhinoceros, socio-ecological analysis, relational database, ecosystem management, episodes detection, taxonomy

## 1. INTRODUCTION

Anthropogenic actions are causing the earth's sixth mass extinction (Ceballos et al., 2015). Such losses occur within *political-ecological processes* wherein sequences of political actions cause social groups to carry out actions that impact ecosystems. Such processes lie at the interface between political science and ecology. The complexity of each of these fields coupled to a pattern of interactions between them can result in highly complex system dynamics. Theoretical understanding of these processes is in its infancy (Bassett and Peimer, 2015). Theory emerges from

efforts to explain observations, and is reinforced or abandoned through the examination of data taken from experiments. Data, then, is critical to the development of theory. As a first step toward structuring data that can support the development of political-ecological theory, a first-of-its-kind database has been developed that is the foundation for creating accessible and searchable databases of political actions that are linked to observations on affected ecosystems. This database is critical to the development of a political-ecological theory of how (a) developed countries interact with developing countries in the management of ecosystems that contain endangered species; (b) groups within developed countries such as conservation-focused *nongovernmental organizations* (NGOs) interact with groups in developing countries such as wildlife management agencies; (c) a country's indigenous people interact with that country's endangered species; and (d) development projects affect a country's endangered species.

Call an ecosystem that contains one or more endangered species, an *at-risk ecosystem*. Although the examples given in this article concern developing countries (Kenya, Tanzania, Uganda, and South Africa), the database developed herein will be of critical importance for understanding the political interactions both within developing and developed countries and between such countries as they affect at-risk ecosystems that span these countries. Specifically, query results from this database can be used to propose and then test a theory of how political actions affect ecosystem management policies; and how the combined effects that these interactions and actions have on what is actually done to conserve or harm an at-risk ecosystem.

Examples of at-risk ecosystems contained in a developed country include the ecosystem of the United States (U.S.) states of Idaho and Montana that contains the reintroduced grey wolf (*Canis lupus*) (Kiasatpour and Whitfield, 2008); and the everglades ecosystem in the U.S. state of Florida that contains the endangered wood stork (*Mycteria americana*). This stork is threatened, in-part, by the invasive Burmese python (*Python molurus bivittatus*) (Dove et al., 2011). And, most challenging, is the Pacific ocean ecosystem that contains the endangered blue whale (*Balaenoptera musculus*) (Haas, 2011, ch. 3). This particular ecosystem spans many developing and developed countries.

Actions are stored in the database in the active voice, e.g., a news article reporting the donation of wildlife monitoring equipment to a wildlife protection agency is entered into the database as a date, actor, and subject-indexed occurrence of the action `donate wildlife monitoring equipment`. Likewise, a news article reporting on the passage of a bill that strengthens wildlife protection laws is entered into the database as a date, actor, and subject-indexed occurrence of the action `strengthen wildlife protection laws`. And a news article reporting on the arrest of members of a wildlife trafficking syndicate is entered into the database as a date, actor, and subject-indexed occurrence of the action `arrest wildlife traffickers`. Entering actions into the database in this way allows it to be queried for particular actions that are taken in a particular time window, by particular actors, and directed toward particular targets (subjects).

This article delivers a multi-faceted breakthrough on how observations on political-ecological actions can be used to develop political-ecological theory. These facets are

1. a first-of-its-kind *relational database* (Churcher, 2016, p. 1–13, 213–231; Coronel and Morris, 2017, p. 72–168; IBM, 2021) to capture the entities that make up political-ecological phenomena, relationships between these entities, and the attributes of these entities;
2. *structured query language* (SQL) scripts (Coronel and Morris, 2017, p. 246–415) for implementing this database technology;
3. the first application of *episode* detection to query results from a political-ecological database; and
4. exemplary use of this relational database to assess political action correlates of east African cheetah habitat loss, and political actions affecting rhino poaching in South Africa.

A political-ecological process or *system* is also called a *socio-ecological* system or a *social-ecological* system (e.g., see Virapongse et al., 2016). The former term is used herein for the following reasons. Ecosystems respond to actions. Hence, the relevant social systems for understanding anthropogenic effects on biodiversity are those that produce actions that affect ecosystems. Political systems are social systems that produce ecosystem management policies and have the capability of using force as necessary to assure these policies are implemented (Moe, 2005; Barthwal and Sah, 2008). In other words, the key characteristic of a political system is that it is capable of exercising coercion. Several groups influence a political system including (a) the country's executive, legislature, and judiciary; (b) international NGOs; (c) organized crime syndicates; and (d) indigenous peoples, racial minorities, and religious minorities. Of these, only group (a) can yield formal governmental power.

But any social system, political or otherwise, is made up of individuals. And many individuals are driven by the need for power (Guinote, 2017). These power-seeking individuals engage in political actions that they believe will increase their sense of power. They do this because political systems can, once controlled, deliver manifestations of the power that these individuals seek. Ultimately, these manifestations of power impact an ecosystem. Hence, a political-ecological system is a complete and precise way to describe the causal chain that starts with individuals wanting power, through political systems that they use to get it, and ultimately to ecosystems that are impacted by its manifestations. On the other hand, the phrase "socio-ecological system" does not convey this power-driven chain of actions.

Data on intentional ecosystem management actions (such as creating a wildlife reserve), and unintentional ones (such as poaching) that is coupled to data on associated ecosystem metrics needs to be easier to use. Doing so would make it easier to assess the effects of management policies on targeted ecosystems (U.S. National Science Foundation, 2014). Rissman and Gillon (2017) believe that links between ecological and social dynamics including feedback loops are needed to inform policy and management and improve both social acceptance and ecological effectiveness of conservation strategies. This view is echoed by Bodin et al. (2014). As Laurila-Pant et al. (2015) develop

a framework for incorporating biodiversity protection into ecosystem management policies, they see a need for large, multi-disciplinary data sets that contain responses to environmental policies and the costs of those policies. And, Leenhardt et al. (2015) see a lack of standardized data on social-ecological systems that links changes in ecological processes to social responses—including attendant feedback loops. These authors believe this scarcity of social-ecological data is limiting the development and validation of social-ecological models.

In response to this general plea for more data structure, Frey and Cox (2015) call for ontologies of political-ecological phenomena to be created. These authors see the building of political-ecological theory being hampered by a lack of data compatibility across different theory-building efforts, and note that data sets collected by different research teams are rarely shared or integrated. They see this data incompatibility challenge as contributing to what they call a "scatter problem:" "...a lack of integration of many research findings into a cohesive set of theoretical instruments that explain how relevant conditions interact to produce success or failure over time." These authors propose the use of ontologies to help reveal observations that are related between different political-ecological databases. Using such data structures would allow integration of different theory building efforts into a comprehensive theory of how politics affects and is affected by the environment. These authors also argue that an ontology supports formal (e.g., relational) database queries via searches based on pre-defined attributes and hence is the preferred way to organize political-ecological data because it structures, unifies, and formalizes the represented knowledge. These characteristics would also allow such knowledge to be reused.

Therefore, an accessible and searchable database of political-ecological actions is seen as a necessary precursor for developing a theory of political-ecological systems and for developing tools to manage such systems. To aid these two development agendas, an *ecosystem management tool* (EMT) has been developed (Haas, 2011, p. 59–78; Haas, 2021) that includes tools for organizing political-ecological data into a relational database. This database implements the *ecosystem management actions taxonomy* (EMAT) of Haas (2011), p. 123–141 and Haas (2018). This taxonomy classifies and indexes political-ecological actions.

A taxonomy is an ontology that represents only hierarchical relationships among its members (American Society for Indexing, 2018). Whereas, in an ontology, there are more types of relationships and these relationships are more specific in their function. Further, information conveyed through indexing in a taxonomy is embedded into the ontology itself (American Society for Indexing, 2018). Taxonomies are also known as *hierarchical ontologies* (Khan and Safyan, 2014). The McKenna/Bell classification system for mammals (Wilson and Reeder, 2005) is an example of a taxonomy. This taxonomy's hierarchy is *Subclass, Infraclass, Supercohort, Cohort, Magnorder, Grandorder, Order,* and *Mirorder*. Another example is the *enterprise ontology* of Dietz (2006). This taxonomy structures those actors and processes that constitute the functioning of an enterprise, e.g., a manufacturing firm.

A *relational database management system* (RDBMS) is the computer system (hardware and software) needed to host the database itself (Mata-Toledo and Cushman, 2000, p. 1). The software component is used to build the database and query it. The present article focuses on the database and software components of a RDBMS. Queries against a relational database are often expressed in SQL. SQL is well-established and used to query about 80% of all databases currently in existence (DB-Engines, 2017). Call a relational database of political-ecological actions developed herein, an *EMAT database*, and the computer system that serves an EMAT database, an *EMAT RDBMS*. Let an observed political action or an observation on an ecosystem metric that has been matched to a member of the EMAT be referred to as an *EMAT action observation*.

A brief tutorial on relational databases appears in **Appendix A**.

Once built, queries against an EMAT database can be used to

1. extract ecosystem actions of a selected type, date range, and country for purposes of assessing an ecosystem's sustainability;
2. form political-ecological data sets to build political-ecological theory through the use of *episode analysis* (described below);
3. construct a data set of EMAT action observations in order to statistically fit the parameters of a simulation model of a political-ecological system (Haas, 2011, p. 161–178);
4. help construct and evaluate ecosystem management policies; and
5. critique, modify, and/or extend the EMAT.

This article delivers as its central contribution, the needed foundation for structuring and using political-ecological data: an EMAT database. Then, as an example of one of the many uses of an EMAT database, episode analysis is developed and used in order to show how it can give insight into a political-ecological system.

Shneider (2009) describes four stages that a new science such as conservation, goes through as it develops. These are (I) the identification of its fundamental objects, phenomena, and language to describe its subject matter; (II) creation of tools for studying these objects and phenomena; (III) discovery of mechanisms that predict observed phenomena; and (IV) broadcast and maintenance of this predictive knowledge. This article delivers a taxonomically-based relational database of political-ecological physical actions, verbal actions, and data. It further develops the concept of an episode of political-ecological actions, and gives a tool for determining if such an episode is causal. These two breakthroughs: an EMAT database and attendant episode analysis give, for the first time, a language to conservation science. This language enables researchers to identify what data needs to be collected, and what a theory of political-ecological systems should be able to explain. In particular, this theory should offer data-verified causal mechanisms that produce the observed, coupled actions of political actors and ecosystem members. This operational triad of objects, phenomena, and language supports the convergence of conservation theories. Therefore, this article makes a fundamental contribution to conservation science

because it completes stage I of a developing science through its EMAT database, and begins stage II through its introduction of episode analysis.

## 2. MATERIALS AND METHODS

### 2.1. EMAT Database Overview

On the political side, an EMAT database holds ecosystem-affecting anthropogenic actions as reported in news articles (hereafter called *stories*). Many of these stories are available online. The story's *source* is the news outlet responsible for the story, e.g., The Huffington Post. Some of these actions can be matched to members of the EMAT. On the ecosystem side, an EMAT database holds actions taken by the ecosystem, and references to ecosystem data sets rather than the data sets themselves. Such data sets are modeled as being observations on the EMAT action `collect data`. This action refers to a set of observations on an ecosystem metric.

Recent efforts to collect biodiversity data have produced large ecological datasets such as the Global Biodiversity Information Facility (GBIF), eBird, iDigBio, and iNaturalist (Heberling et al., 2021). Access to these strictly ecological datasets is free and hence selected subsets of them would be easily added to an EMAT database through the `collect data` EMAT action. Before these datasets can be accessed, however, spatial location of sensitive species observations needs to be generalized so that poachers cannot use the resulting dataset to locate species. There have been several efforts aimed at dealing with this problem (Haas, 2018; Chapman, 2020) although it remains possible (but not likely) that a determined poaching syndicate could hire analysts to reverse engineer species locations.

### 2.2. The EMAT

See Haas (2018) for a detailed description of the EMAT. In brief, the EMAT consists of 632 physical and verbal actions broken into five categories: military, diplomatic, economic, environment, and ecosystem. The fourth category consists of anthropogenic actions directed at the environment, e.g., `clear new land` or `collect data`. The fifth category consists of actions taken by non-anthropogenic actors, e.g., (elephants) `trample crops`—a frequent occurrence in parts of East Africa.

Many of the actions in this taxonomy have been parsed into three *equivalence sets*: A set of semantically equivalent *m*-word verbs, a set of semantically equivalent direct object phrases, and a set of semantically equivalent prepositional phrases. Letting *m* be a positive integer, an *m*-word verb subsumes single-word verbs (either *regular* or *irregular*), and *multi-word* verbs (those that use more than one word to convey their meaning, e.g., "find out") (British Council, 2017). See Aarts (2011) for definitions of direct object phrases, and prepositional phrases.

The EMAT, being an ontology of socio-ecological actions, is an exemplar of what Frey and Cox (2015) see as being needed to advance socio-ecological theory.

## 2.3. Data Acquisition Is Performed Outside of the Database

The task of extracting actions from stories is a cognitive/linguistic data processing activity, namely, shallow parsing coupled to phrase similarity computations. These tasks have little to do with organizing, linking, or querying an existing set of observed political-ecological actions. Therefore, data acquisition is kept separate from an EMAT database. Indeed, there are advantages to separating data acquisition from the building and querying of a relational database. These include the following.

1. Different software systems running on possibly different hardware at different locations can be used to acquire data without the need to transfer such systems to a central location and translate them into a single database language (Nielsen et al., 2013).
2. Computationally intensive data acquisition schemes can be run on special-purpose high performance computing systems without the overhead of an overarching RDBMS.
3. Standard, open source relational database software such as MySQL (Grippa and Kuzmichev, 2021) or Java DB (DB-Engines, 2018) can be used without modification to build and query the database.

See Biermann (2014) for a web-based system that follows this approach of keeping data acquisition separate from database creation, and database querying.
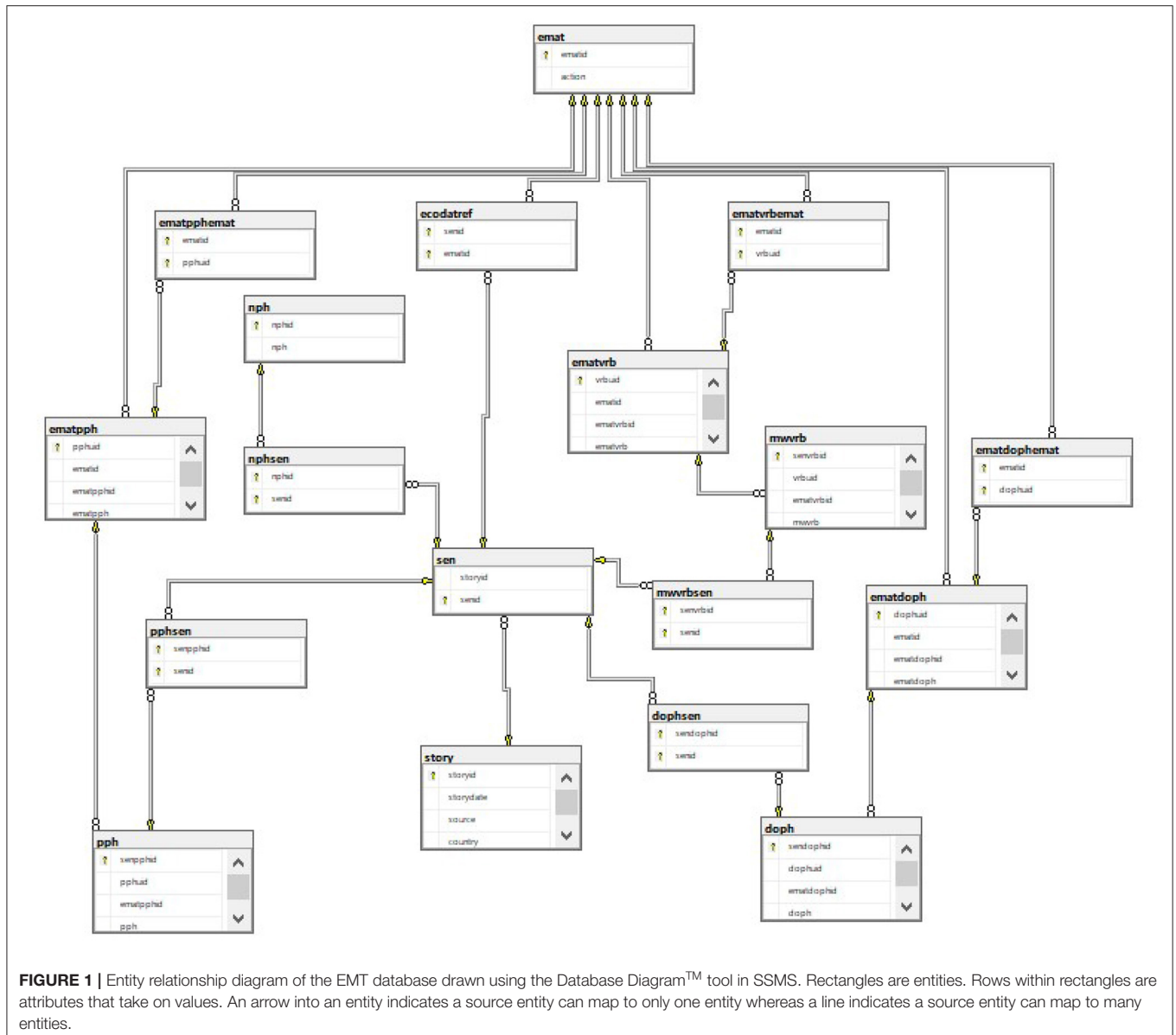
Only sentence components are stored in an EMAT database—not the sentence text itself. This design decision keeps the natural language parsing step (called here, the *parsing preprocessor*) separate from the steps of EMAT database creation and EMAT database querying.

## 2.4. Design of an EMAT Database

The design of a relational database begins with the definition of entities and the relationships between them (Coronel and Morris, 2017, p. 117–168). EMAT database entities are stories, sentences, noun phrases, *m*-word verbs, direct object phrases, prepositional phrases, and EMAT actions. The database schema is hierarchical with stories being at the lowest level followed by sentences; followed by noun phrases, *m*-word verbs, direct object phrases and prepositional phrases—these latter four being at the same level. The highest level in the hierarchy holds the EMAT actions themselves. Actions toward an ecosystem such as `open a wildlife reserve to settlement` are particular EMAT actions. Reactions by the ecosystem to these anthropogenic actions are also EMAT actions. Examples of the latter include the EMAT actions of `trample crops`, and values on landuse over a region contained in a `collect data` EMAT action.

An *entity relationship diagram* illustrates this design (**Figure 1**) wherein for example, an *m*-word verb can map to many sentences and a sentence can map to more than one *m*-word verb: a many-to-many relationship. The requisite junction table in this case is the Figure's `mwvrbsen` table.

The name of the group responsible for an action is contained in that observation's noun phrase entity. The parsing

**FIGURE 1 |** Entity relationship diagram of the EMT database drawn using the Database Diagram™ tool in SSMS. Rectangles are entities. Rows within rectangles are attributes that take on values. An arrow into an entity indicates a source entity can map to only one entity whereas a line indicates a source entity can map to many entities.

preprocessor has assigned this value. Hence, noun phrase entities are exclusively group names, e.g., *Kenya Wildlife Services*, or *rhino poachers*. A group is akin to the social object of an organization as highlighted by Hanneman and Shelton (2011).

Entities that model political action observations have been designed to be at a finer scale than EMAT actions so that new EMAT actions can be added to the EMAT by running learning algorithms that discover new EMAT actions. These algorithms do this by querying an EMAT database for new combinations of group names, *m*-word verbs, direct object phrases, and prepositional phrases (Haas, 2021, **Appendix**). This approach of building a database around parts-of-speech to allow unforseen entity combinations to be discovered is similar to the approach taken by Davies (2005) in his development of a database of the Spanish language.

Entities in the *data set reference* table (table `ecodatref` in **Figure 1**) are observations on the `collect data EMAT` action and have seven attributes: `source`, `species`, `type`, `country`, `region`, `startdate`, and `enddate`. The `source` attribute is either *observation*, or *model*. The `species` attribute indicates the observed species, e.g., *cheetah, rhino*, or *cycad*. The `type` attribute can take on the values of *abundance, presence/absence, capture-recapture, rainfall, NDVI*, and *landuse*. For these latter three values, the species attribute is set to *N/A*. These data set references are preprocessed into one-sentence stories of the form "*group name* collected *type* data on *species* in *region, country* during the period *startdate* to *enddate*." *group name* is the group who collected data, e.g., *Kenya Wildlife Services, SANParks Scientific Services*, or *TerraServer*™.

## 2.5. Software Implementation

The EMAT RDBMS is implemented as an *embedded* Java DB[TM] database (O'Conner, 2006) within the author's **id** software system (see **Appendix B**). This system also contains the parsing preprocessor. An EMAT database is built and queried via **id**'s `rdbms_()` relation. This relation's syntax is

`rdbms_(`*database_name groups_file_name regions_file_name parsed_stories_file_name*

$$line\_1 \ldots line\_m_1 \; endcommand$$
$$[line\_1 \ldots line\_m_2 \; endcommand]$$
$$\vdots$$
$$[line\_1 \ldots line\_m_n \; endcommand]$$

`endquery` *option*`)`

where *line_1 ... line_m_i* is an $m_i$-line SQL query, and *option* is one of `build`, `update`, or `use_existing_database`.

An EMAT database as described in section 2.4 is created when the `build` option is set. This task entails loading the EMAT and each story's sentence components (*m*-word verbs, direct object phrases, and prepositional phrases) into corresponding tables of the Java DB database.

Extracting an EMAT action from a sentence in a story and storing it in an EMAT database is a two-step procedure. The first step, that of retaining a sentence component only if its similarity score is >0.95, is performed within the parsing preprocessor (Haas, 2021, **Appendix**). For a given set of sentence components, the second step consists of computing $\alpha$, the sum of the *m*-word verb similarity score, direct object phrase similarity score, and prepositional phrase similarity score. Then, the associated EMAT action is entered into the database only if $\alpha$ is greater than a threshold value. Some sentences that contain EMAT actions may have no prepositional phrase. For these sentences, $\alpha$ cannot be larger than 1.9. This fact motivates setting the threshold to 1.9 for all EMAT databases reported herein.

## 2.6. Episode Discovery and Causality Testing

An actions history produced by an SQL query against an EMAT database can be used in many ways to aid the development of political-ecological theory. One such way is to characterize the dynamics of political actions by finding repeating temporal patterns of actions, i.e., sequences of actions. Such sequences are called *frequent episodes* in computer science (Ma et al., 2004). Episodes may give some idea of how a group responds to actions of other groups or the ecosystem.

Episodes may also indicate system behaviors that computational theories of political-ecological systems should be able to reproduce. Indeed, a frequently-occurring episode in an observed actions history may be the result of a causal relationship among the groups and ecosystem generating those actions. The presence of such an episode should lead the researcher to apply a test for Granger causality (Budhathoki and Vreeken, 2018) and then examine the political-ecological model for its ability to reproduce the episode should it pass such a test. Call this two-step activity of first computing episodes, and then subjecting them to statistical tests for causality, *episode analysis*.

Python code for conducting the "CUTE" statistical hypothesis test of whether one EMAT action is causing another EMAT action (two time series of binary actions) is available from Budhathoki and Vreeken (2018). To conduct this test, one would first query an EMAT database for only the two actions in question, and then compute the test statistic from the query results. Before the causality of episodes involving more than two EMAT actions can be tested, the CUTE statistic needs to be extended. This extension is straightforward according to Budhathoki and Vreeken (2018).

Observed time series of political-ecological actions is a type of an *observational study* (Rosenbaum, 2002) and hence does not meet the assumptions for a randomized controlled experiment. Granger causality, however, is a widely accepted definition for the effect that one time series has on another. The *compression-based identity test* operationalized by the CUTE statistic provides a means for determining if one sequence of actions is causing another sequence of actions in the sense of Granger causality.

## 2.7. Constructing Complex SQL Queries

Building complex SQL queries to run against an EMAT database can be challenging. Aids for this task exist. For instance, *visual query builders* can help researchers construct complex queries through a graphical user interface that does not require knowledge of SQL. One such tool is the *Query and View Designer*[TM] that is part of Microsoft's *SQL Server Management Studio*[TM] (SSMS) (Microsoft, 2017). Zhang and Yi (1998) and Pankowski (2017) give strategies for developing complex queries.

## 3. RESULTS

## 3.1. Example 2

Three online stories about rhino poaching in South Africa (see **Figures C1–C3**, **Appendix C**) are used to build a database and from that, identify any EMAT action observations they may hold. First, a parsing relation is run in **id** on the file `dbex.txt` to produce a file of parsed stories, `parseddbex.txt`. Then, the desired database query is executed in **id** via the `rdbms_()` relation shown in **Figure 2**. The database is named `polecol`, and is built from the parsed stories contained in the file `parseddbex.txt` using group names contained in the file `sarhinogroups.dat`, and region names contained in the file `sarhinorgns.dat`. This run produces a list of EMAT action observations (**Table 1**).

The SQL query of **Figure 2** consists of selecting only unique (`distinct`) entries from a list formed by joining records from the *m*-word verb table and the direct object phrase table that match on their story source, sentence source, and EMAT action attributes. This complex query appears simple when visualized (**Figure 3**).

**Figure 2** is central to this article's main point: The **id** language relation shown therein is the only interface to an EMAT database that a researcher needs for purposes of extracting data. The idea is to embed the political-ecological system being observed into one software system, here, **id**, along with the formal database language, SQL to run queries against it.

```
report prepare_data
    rdbms(polecol sarhinogroups.dat sarhinorgns.dat parseddbex.txt
        select distinct v.storyid, v.senid, v.mwvrb, d.doph, v.action
        /* List all EMAT actions found in the parsed sentences of dbex.txt. */
        from (
            /* mwvrb subquery. */
            select subvs.storyid, subvs.senid, subv.mwvrb, action
            from (select * from mwvrb, ematvrbemat
                join emat on emat.ematid = ematvrbemat.ematid
                join ematvrb on ematvrb.vrbuid = ematvrbemat.vrbuid
                where mwvrb.mwvrb = ematvrb.ematvrb) subv
            join (select sen.storyid, mwvrbsen.senid, mwvrb from mwvrbsen
                join mwvrb on mwvrb.senvrbid = mwvrbsen.senvrbid
                join sen on sen.senid = mwvrbsen.senid) subvs
            on subv.mwvrb = subvs.mwvrb) v
        join (
            /* doph subquery. */
            select subds.storyid, subds.senid, subd.doph, action
            from (select * from doph, ematdophemat
                join emat on emat.ematid = ematdophemat.ematid
                join ematdoph on ematdoph.dophuid = ematdophemat.dophuid
                where doph.doph = ematdoph.ematdoph) subd
            join (select sen.storyid, dophsen.senid, doph from dophsen
                join doph on doph.sendophid = dophsen.sendophid
                join sen on sen.senid = dophsen.senid) subds
            on subd.doph = subds.doph) d
        on v.storyid = d.storyid and v.senid = d.senid and v.action = d.action
        endquery
        build)
```

**FIGURE 2 |** An `rdbms_()` relation to build the `polecol` EMAT database of Example 2 and query it for EMAT actions.

**TABLE 1 |** Sentence components found in Example 2's three stories by the EMAT action extraction algorithm along with associated EMAT action observations.

| Story and sentence | *m*-word verb | Direct object phrase | Prepositional phrase | Observed EMAT action |
|---|---|---|---|---|
| 1, 1 | Discovery | Of dead rhinos | | Sell a few rhino horns |
| 1, 2 | Arrests | Of alleged poachers | | Arrest rhino poachers |
| 1, 12 | Intercepted | Consignment | Of rhino horn | Seize rhino horn |
| 2, 18 | Combats | Rhino poaching | With new regional | Increase antipoaching |
| | | | Court | Enforcement |
| 3, 25 | Shot dead | Rhino | | Sell a few rhino horns |

## 3.2. Two Operational EMAT Databases

Two EMAT databases capable of supporting theory development have been constructed. The first focuses on actions that affect the cheetah population in East Africa (Kenya, Tanzania, and Uganda) (Haas, 2019a). The second focuses on the rhino population in South Africa (Haas, 2019b). Those stories pertaining to East African cheetah that were successfully parsed are contained in the database named east-af-cheetah. Likewise, those stories pertaining to South African rhinos that were successfully parsed are contained in the database named south-af-rhinos. The set of raw HTML files that contain these stories along with statistics that measure the EMAT action extraction algorithm's productivity are shown in **Table 2**.

These two applications have been chosen in order to highlight the ability of the EMAT database technology described herein to capture fundamentally different political mechanisms and actions that affect and are affected by the actions of an at-risk ecosystem.

### 3.2.1. Action Extraction

**Figures 4**, **5** show the time series of the actions extracted by the EMAT action extraction algorithm of section 2.5 applied to the East African stories, and South African stories, respectively (see **Table 3**). Cheetah abundance data is from Durant et al. (2017), IUCN/SSC (2007), and TMAP (2008). Note the prominence of reports of human-wildlife conflict in the East African press, and the prevalence of rhino poaching reports (sell a few rhino horns) in the South African press. These two Figures highlight the dynamic nature of this data and also the temporal linkages between anthropogenic actions and ecosystem reactions to them.

**TABLE 2 |** Raw HTML story files available on the East African cheetah ecosystem management tool website (Haas, 2019a), and South African rhino ecosystem management tool website (Haas, 2019b).

| Phrase used for web search | File name | Number of online stories | Number of stories parsed |
|---|---|---|---|
| **Stories associated with East African cheetah** | | | |
| | Efstories-manual.txt | 605 | 279 |
| | Efstories-alerts.txt | 1,526 | 359 |
| "Cheetah" | Efstories-script09-13.txt | 695 | 243 |
| | EFSTORIES-script13-14.txt | 3,101 | 1,438 |
| | Efstories-script14-15.txt | 1,856 | 1,130 |
| | Efstories-script15-16.txt | 6,519 | 2,053 |
| | Efstories-alerts16-18.txt | 3,519 | 1,339 |
| | Efstories-alerts18-19.txt | 3,644 | 958 |
| Totals | | 21,465 | 7,799 |
| **Stories associated with South African rhinos** | | | |
| | Kruger080313.txt | 1,184 | 452 |
| "Kruger rhino poaching" | Kruger103114.txt | 986 | 265 |
| | Kruger120715.txt | 1,081 | 339 |
| | Kruger5417.txt | 473 | 123 |
| Totals | | 3,724 | 1,179 |

*Stories are scraped from online news aggregators using the given search phrase. An article is successfully parsed if the algorithm can detect a date, country, and at least one action. Performance measures on the number of stories successfully parsed demonstrate the parsing preprocessor's productivity.*

### 3.2.2. Episodes

Because an episode is a sequence of actions, each action in a particular episode has a *position*. The Position Pair Set (PPS) algorithm given in Ma et al. (2004) is used to discover such episodes in the east-af-cheetah database (**Table 3**). The first East African cheetah episode is reflective of the attention that human-wildlife conflict receives in the East African press. The first episode in the South African rhino actions history reflects the nearly constant repetition of rhino poaching reports in the South African press.

### 3.2.3. Cheetah Habitat Destruction

The East African cheetah EMAT database can aid research into the political antecedents that drive reductions in cheetah habitat. Ideal habitat for cheetah is nearly open plains with a shrub cover of about 40%, and a few kopjes (rocky outcroppings) (Broekhuis, 2017). Such habitat can be lost through several politically driven or sanctioned actions including the degazzetting of wildlife reserve land, clearing land for farms or ranches, urban sprawl, and the construction of new roads (Learn, 2017). The four SQL commands and one SQL query used to extract all political actions that negatively impact cheetah habitat are shown in **Figure 6**. The resulting actions history appears in **Figure 7**.

**Figure 6** is complex SQL query that is composed of two separate SQL commands: The creation of two temporary tables (v and d) followed by a join operation on them. Temporary table v collects *m*-word verbs associated with cheetah habitat-damaging political actions, and temporary table d collects those

direct object phrases that are associated with these actions. Entries in these two tables are then selected that match on their story source, sentence source, and action attributes.

There are 11,620 stories in the east-af-cheetah database. This SQL query requires 45 min on a PC running at 3.0 GHz. As can be seen by the plot, these actions are regular and on-going. Two frequently-occurring episodes that involve the habitat-destroying action open reserve to settlement are shown in **Table 4**. These episodes suggest that this action may be a consequence of the decision to invest in tourism but an antecedent to wildlife crime. Having discovered these temporal associations through episode extraction, statistical testing might reveal that such sequences are actually causal.

## 4. DISCUSSION

### 4.1. Related Work

#### 4.1.1. Relational Databases of Socio-Ecological Data

A search was conducted of the peer-reviewed literature for reports on the development of relational databases of temporally-indexed observations on social actions that are linked to observations on ecosystem metrics. None were found. Such data is in contrast to databases of case studies of social-ecological systems such as SESMAD (Cox, 2014). What was found, however, were three non-relational data sets that are somewhat similar to what was sought. In the most relevant of these (Xie et al., 2019), the authors construct a socio-ecological data set by coupling socioeconomic statistics to climate data. Then, they use their data set to discover feedback loops between grassland productivity and human actions. Two less-similar data sets are (a) a data set of one-time observations on social and ecological variables within the Amazon basin (Lima et al., 2016); and (b) a one-time data set of descriptive (non-event) social-ecological data on pastoral systems in Mongolia (Laituri et al., 2015).

#### 4.1.2. Construction of Social Theory

Davies (2005) describes a landmark relational database of a corpora of medieval and modern Spanish texts. This author argues that the relational database structure allows investigators to query the database for output to test new theories about how that language evolved. In the field of law, Ribary (2020) presents a relational database of Roman law and argues that queries against it will help tie together many investigations into how legal systems in the ancient world worked. And in the related field of criminal justice, there are several relational databases of terrorists and terrorism events (Bowie, 2017).

The potential for relational databases to aid the development of social science theory in general is discussed in Hanneman and Shelton (2011). These authors address the question of how one would mine several different social process databases including (a) the periodical literature, (b) business directories, and (c) international trade digests. These authors see such potential because of the critical role in social science theory development already played by the databases *Sociological Abstracts* (ProQuest, 2010), and *Web of Science* [Institute for Scientific Information (ISI), 2010]. Similarities are drawn between social processes and the database concepts of an object (entity), and relations
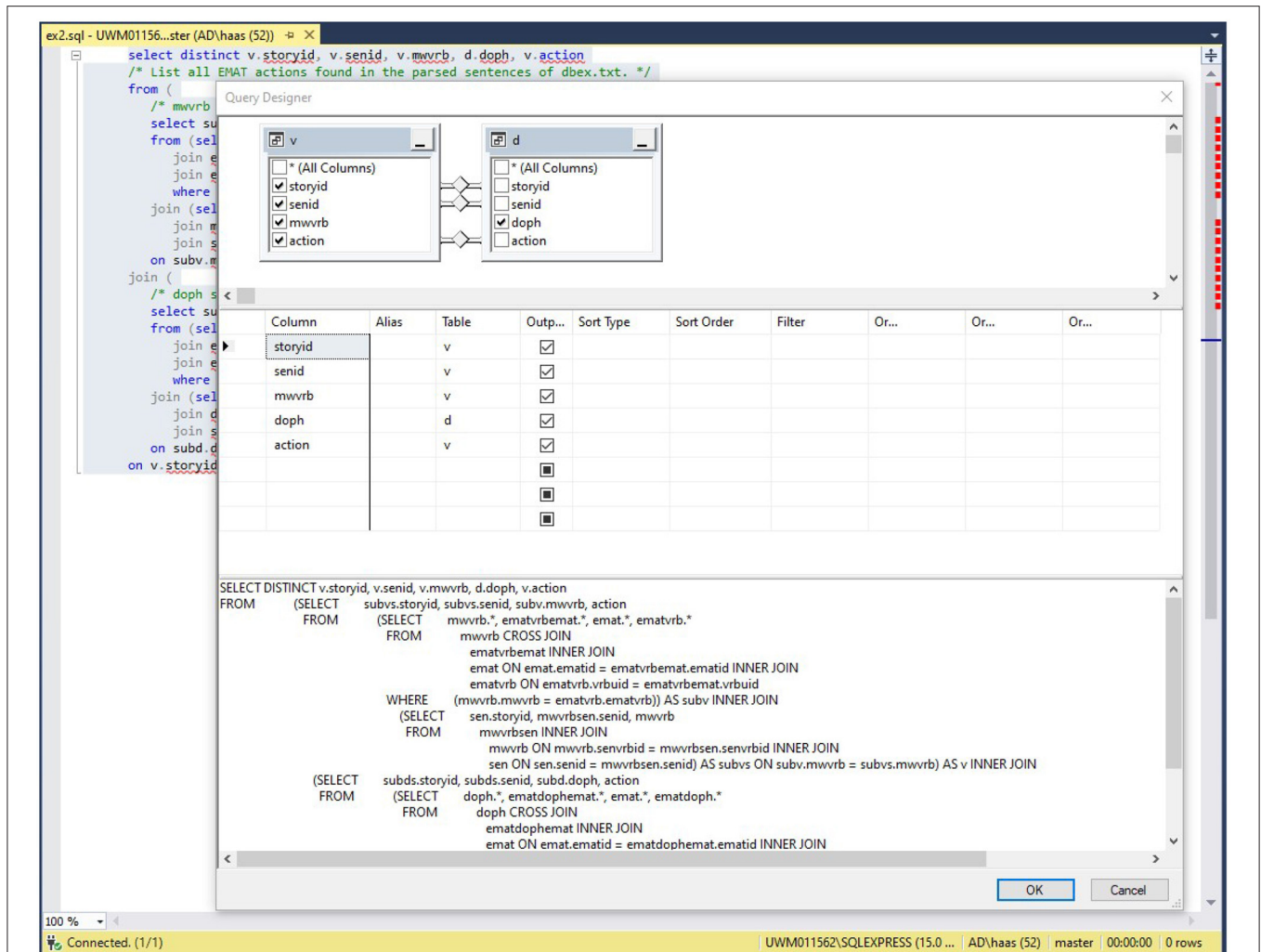
**FIGURE 3 |** Query diagram of the SQL query of **Figure 2**. The diagram appears in the *Diagram pane* of the Query and View Designer tool. This pane allows query design through drag-and-drop mouse operations. The middle screen is the *Criteria pane* and allows query design through spreadsheet-type entries. The bottom screen is the *SQL pane* and contains the parsed version of the hand-written SQL code that appears in the background window. The red bars on the right-hand side of this window indicate the beginning and content of SQL `select` commands.

between entities. A distinction is made between social objects that are not people such as events and organizations that do, however, possess attributes and agency. These authors point out that the development of social theory is aided when these objects can be classified by their similarities, i.e., similar values on their relational database attributes. And, they argue that social transactions can be modeled as entities connected through relational database linkages. Doing so would enable theories of information flow between social groups to be postulated and tested.

One could conclude from these remarks that building a relational database of a temporal, dynamic social process can contribute to the understanding of that process's dynamics. And further, having such a database would allow the use of the powerful analytic tool, SQL to extract theory-motivated subsets of data from it.

## 4.2. Peripheral Work
### 4.2.1. Flat Files?

Why not use a simple collection of text-based computer files to hold political-ecological data? Managing data in this manner is commonly referred to as maintaining a *flat file database* (Database Management System, 2017). If built within a relational database software system, a flat file database is a single-table relational database. If instead, the database is implemented as a set of computer files outside of a relational database software system, the resultant database will suffer from several deficiencies as follows.

1. There is no data model. Hence, a common ontology of political-ecological actions cannot be developed and shared among researchers.
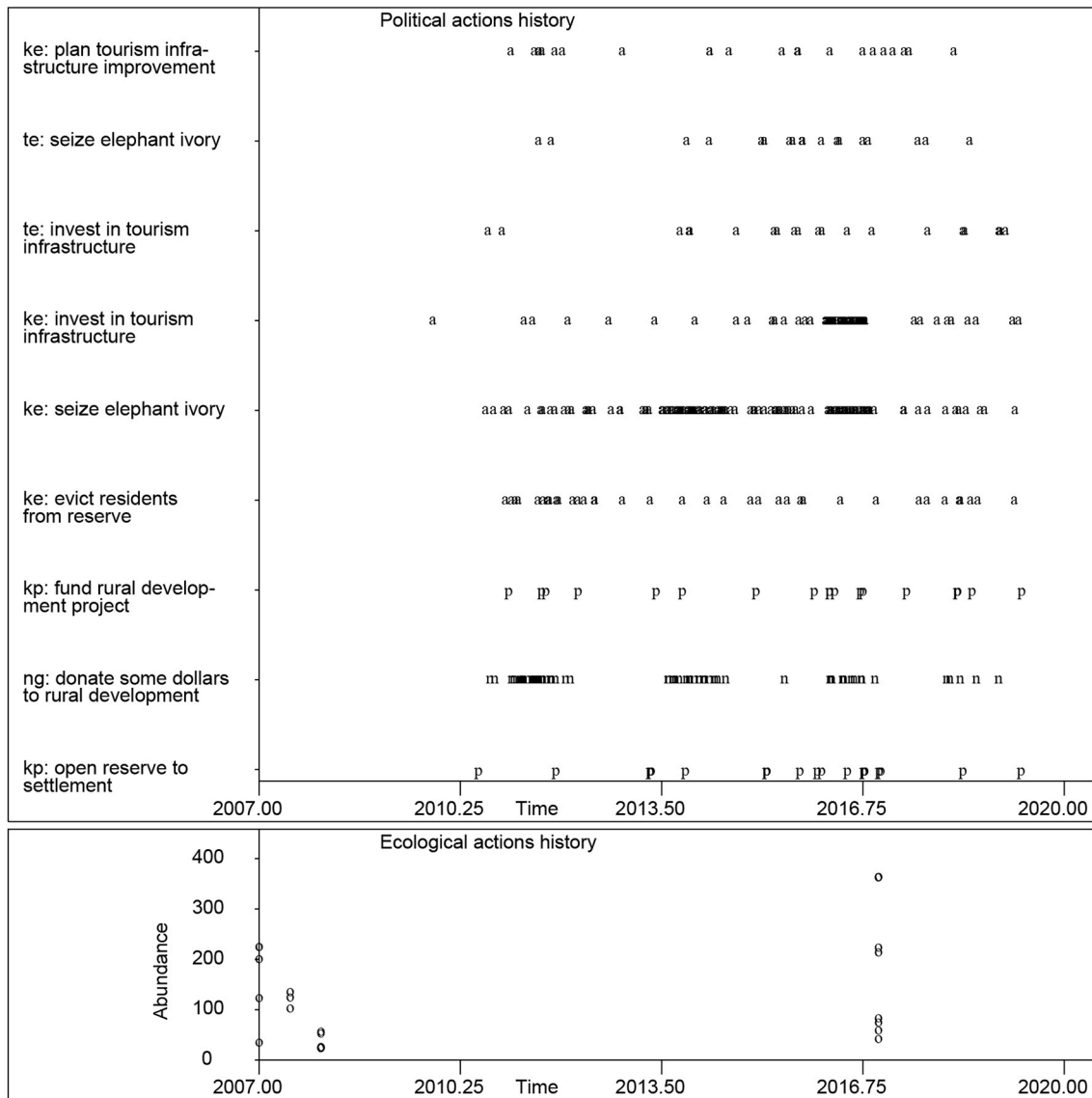
**FIGURE 4 |** Observed actions history from East African online news stories for the period from January 2007 to June 2019. The presidential office of Kenya, Tanzania, and Uganda are designated by `kp`, `tp`, and `up`, respectively. Similarly, the environmental/wildlife protection agencies are designated with `ke`, `te`, and `ue`; non-pastoralist, rural residents with `kr`, `tr`, and `ur`; and pastoralists with `ka`, `ta`, and `ua`. The group of conservation NGOs who have operations in at least one of these countries is represented by `ng`. The plotting symbol `p` indicates an action taken by a presidential office, `a` an action taken by an EPA, `r` an action taken by rural residents, `s` an action taken by pastoralists, `d` an action taken by a developer, `s` an action taken by pastoralists, `t` an action taken by tourists, `l` an action taken by a legislature, `j` an action taken by a judge, and `n` an action taken by an NGO. Only frequent out-combinations are shown. The bottom plot is observed cheetah abundance.

2. Only pre-programmed queries can be processed. This is called *program-data dependency*. Unanticipated queries (*ad-hoc queries*) can only be made by writing and then executing custom application programs. But, because ecosystem management is an emerging discipline, it is not possible to know in advance what forms of political-ecological data that modelers and policymakers will require. Hence, *ad-hoc queries* will be the norm rather than the exception. In addition, users of a flat file database need to be skilled in writing computer programs. In light of the above-mentioned need for the development of ecosystem management tools aimed at creating sustainable ecosystems, researchers, and policymakers from a wider range of backgrounds will need to become involved. Only a subset of these individuals can be expected to have such requisite programming skills. A lack of such skills then, has the potential of making data access a critical bottleneck to more effective management of at-risk ecosystems.
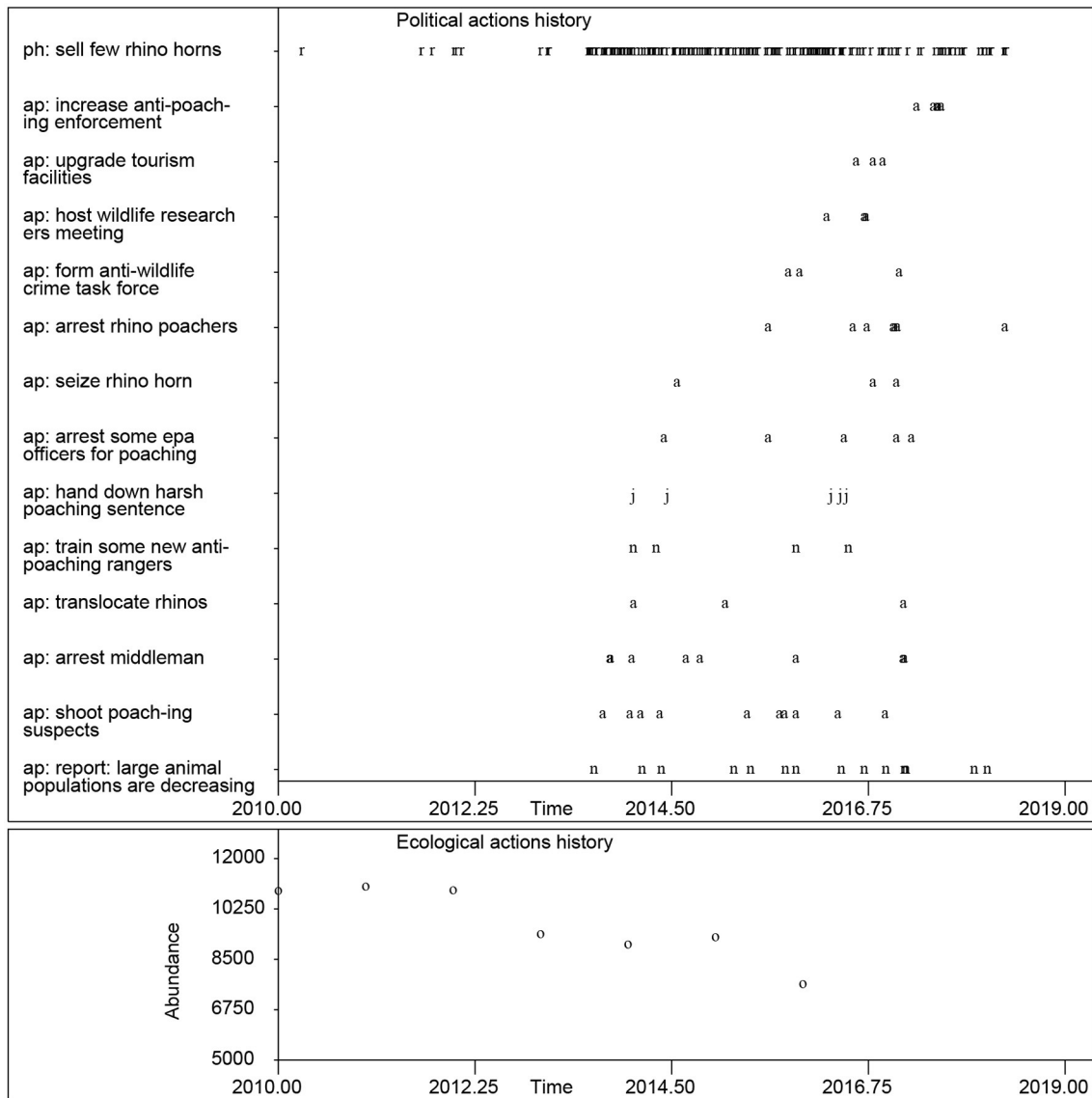
**FIGURE 5 |** Observed actions history from South African online news stories for the period January 2010–June 2019. Rhino poachers are designated by `ph`. South African rangers and administrators engaged in anti-poaching activities are designated by `ap`. Rhino abundance data is from Haas and Ferreira (2018). See **Figure 4** for the plotting symbol legend.

3. Data isolation can occur when distinct collections of data that are instances of the same entities are held in different files as there is no mechanism to recognize such relationships.

4. If data on two or more entities is redundant, data inconsistencies can emerge. But a flat file system has no mechanism to guard against such redundant data.

5. Data corruption can occur through multiple users accessing the same records of one or more of the files. These events are called *concurrent access anomalies.*

6. No checks are made on data integrity, e.g., a restriction on the range of an attribute.

7. No security protocols are implemented to control who has access to the data.

See CareerBless (2018) for further details of these deficiencies. Wesley (2000) gives a step-by-step tutorial on how a flat file database can be improved by migrating it to a relational database.

### 4.2.2. Merging Heterogeneous Databases

A topic related to schema development is the merging of heterogeneous databases (e.g., Karasneh et al., 2009). These authors introduce a *relations schemas matcher* algorithm that produces a measure of similarity between the names of tables in two different databases. This syntactic phrase similarity measure is related to the *SIM(phrase_1, phrase_2)* algorithm given in Haas (2021) (**Appendix**). The focus of the present work, however, is

**TABLE 3 |** A selection of frequent, multi-action episodes in the East African cheetah, and South African rhino actions history data sets.

| Episode | Frequency | Position in sequence | Action |
|---|---|---|---|
| | | **East African cheetah** | |
| 1 | 15 | 1 | Elephants_trample_crops |
| | | 2 | Hippo_attacks_some_cattle |
| 3 | 11 | 1 | Donate_equipment_for_anti-poaching_enforcement |
| | | 2 | Evict_residents_from_reserve |
| 2 | 7 | 1 | Seize_elephant_ivory |
| | | 2 | Donate_equipment_for_anti-poaching_enforcement |
| | | 3 | Elephants_trample_crops |
| | | **South African rhinos** | |
| 1 | 151 | 1 | Sell_a_few_rhino_horns |
| | | 2 | Sell_a_few_rhino_horns |
| 2 | 9 | 1 | Report:_large_animal_populations_are_decreasing |
| | | 2 | Sell_a_few_rhino_horns |
| 3 | 9 | 1 | Sell_a_few_rhino_horns |
| | | 2 | Sell_a_few_rhino_horns |
| | | 3 | Shoot_poaching_suspects |

*Episodes are found with the PPS algorithm of Ma et al. (2004).*

on the development of a new relational database of political-ecological actions that have been taken from original, non-database sources.

## 4.3. Constructing Ecosystem Management Tools

In addition to its role in developing political-ecological theory, an EMAT database is engineered to support existing and future ecosystem management decision support tools. How EMAT databases are, and should be employed in these two support roles is described in the following sections.

### 4.3.1. Supporting Existing Tools

Efforts are underway to build political-ecological system models that represent the effect of different ecosystem management policies. If such models were statistically fitted to political-ecological data, their use in assessing proposed ecosystem management options would be more credible (Haas, 2020). For example, Haas and Ferreira (2018) develop an agent-based model of the political-ecological system that contains the South African rhino population. As these authors state: "One part of this model is a submodel of poachers interacting with the rhino population. Combinations of antipoaching initiatives and economic opportunities are evaluated as to their probable effectiveness at changing local people's inclination to poach rhinos and the consequent effect on the rhino population." These authors first fit their model to a political-ecological data set composed of coupled observations on poaching actions, antipoaching actions, and a `collect data` set of rhino abundance observations. Then, a particular management policy is evaluated by running the fitted model under this policy and examining its effect on rhino abundance.

As another example, Miyasaka et al. (2017) develop an agent-based, social-ecological model of land use by farmers in Inner Mongolia. They use this model to study the impact of different land management policies on the impacted dryland ecosystem. Their model is spatially-explicit, incorporates a learning mechanism, and enacts two-way interaction with the ecosystem (anthropogenic effects on the ecosystem and feedback effects by the ecosystem on the farmers). Their model also allows for time lags. They empirically calibrate the farmer submodel with survey data, and the ecosystem submodel with biophysical measurements. If their social-ecological data were entered into an EMAT RDBMS, model calibration exercises could be repeated in a transparent way as new data became available. Or, more impactfully, such an RDBMS could support the statistical fitting of the model's parameters with (say) maximum likelihood. This RDBMS would make the many database queries needed to accomplish this model-wide statistical estimation exercise easier to organize, perform, and communicate.

### 4.3.2. Supporting Future Tools

Kupschus et al. (2016) call for *integrated monitoring programmes* to support the ecosystem approach to conservation management. These authors argue that a monitoring program needs to provide the means to test causal relationships between anthropogenic actions and ecosystem responses. Such monitoring would produce large amounts of political-ecological data. This data would need to be housed in an accessible way. In a similar vein, Ascough et al. (2008) identify challenges to assessing models used to support environmental and ecological decision-making. These authors see a need for a "holistic, integrated uncertainty framework" in order to comprehensively incorporate uncertainty into environmental decision making. They propose a web-based system as one way to implement such a framework.

```
report prepare_data
    rdbms(east-af-cheetah efgroups.dat efregions.dat efprsd.dat
    declare global temporary table v (storyid int, senid int,
        mwvrb varchar(30000), action varchar(80)) not logged
    endcommand
    insert into session.v
    SELECT distinct subvs.storyid, subvs.senid, subv.mwvrb, action
    FROM (SELECT mwvrb.vrbuid, mwvrb.mwvrb, emat.action
            FROM mwvrb, ematvrbemat
            JOIN emat ON emat.ematid = ematvrbemat.ematid
            JOIN ematvrb ON ematvrb.vrbuid = ematvrbemat.vrbuid
            WHERE mwvrb.mwvrb = ematvrb.ematvrb
            AND (emat.action IN ('open_reserve_to_settlement',
                    'reduce_game_reserve_size',
                    'sell_grazing_land_to_individuals',
                    'destroy_lake_from_farming_and_industry',
                    'block_migration_routes_with_settlements',
                    'build_road_through_NP', 'clear_new_land'))
            ) subv
    JOIN (SELECT sen.storyid, mwvrbsen.senid, mwvrb
            FROM mwvrbsen
            JOIN mwvrb ON mwvrb.senvrbid = mwvrbsen.senvrbid
            JOIN sen ON sen.senid = mwvrbsen.senid) subvs
    ON subv.mwvrb = subvs.mwvrb
    endcommand
    declare global temporary table d (storyid int, senid int,
        doph varchar(30000), action varchar(80)) not logged
    endcommand
    insert into session.d
    SELECT distinct subds.storyid, subds.senid, subd.doph, action
    FROM (SELECT doph.dophuid, doph.doph, emat.action
            FROM doph, ematdophemat
            JOIN emat ON emat.ematid = ematdophemat.ematid
            JOIN ematdoph ON ematdoph.dophuid = ematdophemat.dophuid
            WHERE doph.doph = ematdoph.ematdoph
            AND (emat.action IN ('open_reserve_to_settlement',
                    'reduce_game_reserve_size',
                    'sell_grazing_land_to_individuals',
                    'destroy_lake_from_farming_and_industry',
                    'block_migration_routes_with_settlements',
                    'build_road_through_NP', 'clear_new_land'))
            ) subd
    JOIN (SELECT sen.storyid, dophsen.senid, doph
            FROM dophsen
            JOIN doph ON doph.sendophid = dophsen.sendophid
            JOIN sen ON sen.senid = dophsen.senid) subds
    ON subd.doph = subds.doph
    endcommand
    SELECT story.storydate, v.action
    from session.v v
    JOIN session.d d
    ON v.storyid = d.storyid AND v.senid = d.senid AND v.action = d.action
    JOIN story ON story.storyid = v.storyid
    endquery
    build)
```

**FIGURE 6 |** An `rdbms_()` relation to build temporary tables in the `east-af-cheetah` EMAT database and then query them for habitat-destroying EMAT actions.

These authors envisage a system that contains two databases: One holding ecosystem observations, e.g., geospatial data, and one holding output from models of ecosystem function. These databases would be used to compare the quality and precision of different ecosystem models. Quality would consist of side-by-side comparisons of reliability and validity; and precision would consist of side-by-side comparisons of values on uncertainty measures, e.g., prediction intervals of a biodiversity index across time. These comparisons would be used by both scientists and policymakers to select models to use to inform what actions to take to manage an at-risk ecosystem. The two constituent databases would contain large numbers of interconnected variables and many multivariate observations on them. Although the authors are silent on the nature of these databases, their vision for how data would be used by the system points to a need for databases of ecosystem observations that are as comprehensive and accessible as possible. A single EMAT RDBMS would be an accessible portal capable of holding enough variables and observations to achieve such comprehensiveness. This would be done by populating the database with pairs of observations on the `collect data` EMAT action—one member of a pair having the `source` attribute set to *observed*, and the other with `source` set to *model*.

## 4.4. Conclusions

This article has demonstrated that data on taxonomy-indexed political-ecological actions can be organized into an EMAT RDBMS implemented within a single, stand-alone Java program.
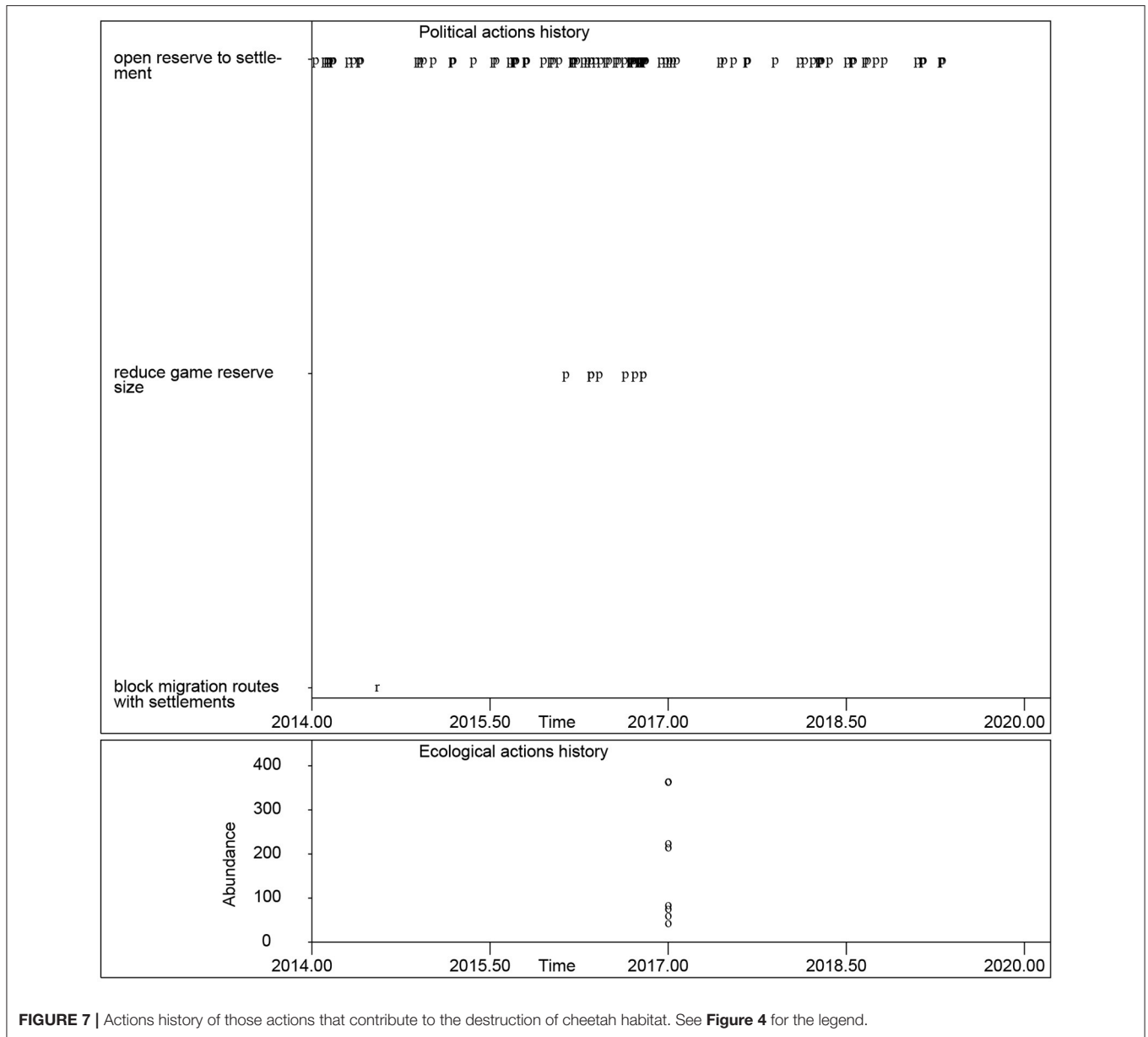
**FIGURE 7 |** Actions history of those actions that contribute to the destruction of cheetah habitat. See **Figure 4** for the legend.

**TABLE 4 |** Episodes in the East African cheetah database that contain the habitat-destroying EMAT action of `open reserve to settlement`.

| Episode | Frequency | Position in sequence | Action |
|---------|-----------|----------------------|--------|
| 1 | 6 | 1 | Invest_in_tourism_infrastructure |
|   |   | 2 | Open_reserve_to_settlement |
| 2 | 8 | 1 | Open_reserve_to_settlement |
|   |   | 2 | Seize_elephant_ivory |

The exercise of building a relational database for political-ecological data illustrates how the requirement that database tables be linked forces the researcher to hypothesize about how different entities and components of political-ecological actions might be related to each other. At this early stage in our understanding of how groups and ecosystems react to different policies, such an exercise can only be beneficial.

Given the continuing decline in global biodiversity (Ceballos et al., 2015), there is an immediate need for a larger group of people to become involved in ecosystem management. Social scientists need to work with ecologists, business leaders, and government officials to quickly find effective policies to manage the interface between humans and ecosystems. The database technology described herein, if implemented on data taken on the planet's most at-risk ecosystems, would help to bridge these gaps. In order to do this, however, these databases will need to be connected to persistent streams of political-ecological data.

## 4.5. Next Steps
### 4.5.1. Shortcomings
EMAT action equivalence sets are manually developed and have been completed for only a small subset of the actions in the EMAT. This step in the parsing preprocessor needs to be automated with current parsing algorithms that reference a large corpora of international stories. Complex queries that are run against large EMAT databases can be computationally expensive. These databases and queries need to be in a RDBMS that is running on a high performance computing system. The porting and tuning required to achieve acceptable runtimes on such systems is a nontrivial task and further, not all researchers will have access to such systems. Maintaining an EMAT database that is connected to online story feeds requires an investment in hardware and trained staff. This resource requirement may keep many researchers from developing their own EMAT databases.

### 4.5.2. The Future
The work of the present article suggests two research directions. The first is to compile a list of all current political-ecological models and then update the EMAT database schema so that its entities, attributes, and linkages represent the observable variables of as many of these models as possible. Doing so would allow a common language of political-ecological theory to evolve as called for by Frey and Cox (2015). Second, an EMAT database needs to be built for each ecosystem that hosts species who are at risk of extinction. These databases would then be used to find management policies that conserve them.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

TH developed the theory, wrote all computer code, created all figures and tables, and wrote the article.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcosc.2021.707088/full#supplementary-material

## REFERENCES

Aarts, B. (2011). *Oxford Modern English Grammar*. Oxford University Press.

American Society for Indexing (2018). *Taxonomies & Controlled Vocabularies Special Interest Group*. Available online at: www.taxonomies-sig.org/about.htm (accessed January 30, 2020).

Ascough, J. C. II, Maier, H. R., Ravalico, J. K., and Strudley, M. W. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecol. Model.* 219, 383–399. doi: 10.1016/j.ecolmodel.2008.07.015

Barthwal, C. P., and Sah, B. L. (2008). Role of governmental agencies in policy implementation. *Indian J. Polit. Sci.* 69, 457–472. Available online at: http://www.jstor.org/stable/41856437

Bassett, T. J., and Peimer, A. W. (2015). Political ecological perspectives on socioecological relations. *Nat. Sci. Soc.* 23, 157–165. doi: 10.1051/nss/2015029

Biermann, M. (2014). A simple versatile solution for collecting multidimensional clinical data based on the CakePHP web application framework. *Comput. Methods Prog. Biomed.* 114, 70–79. doi: 10.1016/j.cmpb.2014.01.007

Bodin, Ö., Crona, B., Thyresson, M., Golz, A.-L., and Tengö, M. (2014). Conservation success as a function of good alignment of social and ecological structures and processes. *Conserv. Biol.* 28, 1371–1379. doi: 10.1111/cobi.12306

Bowie, N. G. (2017). Terrorism events data: an inventory of databases and data sets, 1968-2017. *Perspect. Terror.* 11, 50–72.

British Council (2017). *Multi-Word Verbs*. Available online at: https://learnenglish.britishcouncil.org/en/quick-grammar/multi-word-verbs (accessed January 30, 2020).

Broekhuis, F. (2017). *Habitat selection patterns of cheetahs Acinonyx Jubatus in the Serengeti, Tanzania* [Masters thesis]. Institute of Zoology and the Royal Veterinary College, London, United Kingdom. Available online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.6036&rep=rep1&type=pdf (accessed January 30, 2020).

Budhathoki, K., and Vreeken, J. (2018). "Causal inference on event sequences," in *Proceedings of the 2018 SIAM International Conference on Data Mining* (San Diego, CA).

CareerBless (2018). *Traditional Approach for Data Storage and the Need for DBMS*. Available online at: www.careerbless.com/db/rdbms/c1/traditionalApproach.php (accessed January 30, 2020).

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., and Palmer, T. M. (2015). Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci. Adv.* 1:e1400253. doi: 10.1126/sciadv.1400253

Chapman, A. D. (2020). *Current Best Practices for Generalizing Sensitive Species Occurrence Data*. Copenhagen: GBIF Secretariat.

Churcher, C. (2016). *Beginning SQL Queries*. New York, NY: Apress Media, LLC.

Coronel, C., and Morris, S. (2017). *Database Systems: Design, Implementation, & Management, 12th Edn*. Boston, MA: Cengage Learning.

Cox, M. (2014). Understanding large social-ecological systems: introducing the SESMAD project. *Int. J. Commons*. 8, 265–276. doi: 10.18352/ijc.406

Database Management System (2017). *Flat File Systems and Their Drawbacks (Traditional File System)*. Available online at: www.databasemanagementsystemsz.com/flat-file-systems-drawbacks-traditional-file-system-database-management-system/ (accessed January 30, 2020).

Davies, M. (2005). The advantage of using relational databases for large corpora. *Int. J. Corpus Linguist.* 10, 307–334. doi: 10.1075/IJCL.10.3.02DAV

DB-Engines (2017). *Ranking Scores Per Category in Percent*. Available online at: https://db-engines.com/en/ranking_categories (accessed January 30, 2020).

DB-Engines (2018). *System Properties Comparison Derby vs. MySQL*. Available online at: https://db-engines.com/en/system/Derby;MySQL (accessed January 30, 2020).

Dietz, J. L. G. (2006). *Enterprise Ontology*. Springer-Verlag.

Dove, C. J., Snow, R. W., Rochford, M. R., and Mazzotti, F. J. (2011). Birds consumed by the invasive Burmese python (Python molurus bivittatus) in Everglades National Park, Florida, USA. *Wilson J. Ornithol.* 123, 126–131. doi: 10.1676/10-092.1

Durant, S. M., Mitchell, N., Groom, R., Pettorelli, N., Ipavec, A., Jacobson, A. P., et al. (2017). The global decline of cheetah *Acinonyx jubatus* and what it means for conservation. *Proc. Natl. Acad. Sci. U.S.A.* 114, 528–533. doi: 10.1073/pnas.1611122114

Frey, U. J., and Cox, M. (2015). Building a diagnostic ontology of social-ecological systems. *Int. J. Commons* 9, 595–618. doi: 10.18352/IJC.505

Grippa, V. M., and Kuzmichev, S. (2021). *Learning MySQL: Get a Handle on Your Data, 2nd Edn*. New York, NY: O'Reilly Media.

Guinote, A. (2017). How power affects people: activating, wanting, and goal seeking. *Annu. Rev. Psychol.* 68, 353–381. doi: 10.1146/annurev-psych-010416-044153

Haas, T. (2020). Developing political-ecological theory: the need for many-task computing. *PLoS ONE* 15:e0226861. doi: 10.1371/journal.pone.0226861

Haas, T. C. (2011). *Improving Natural Resource Management: Ecological and Political Models.* Milwaukee, WI: Wiley-Blackwell.

Haas, T. C. (2018). Automatic acquisition and sustainable use of political-ecological data. *Data Sci.* 17:17. doi: 10.5334/dsj-2018-017

Haas, T. C. (2019a). *Cheetah Ecosystem Management Tool, Online Resource.* Available online at: https://sites.uwm.edu/haas/home/cheetah_emt/ (accessed January 30, 2020).

Haas, T. C. (2019b). *Rhino Ecosystem Management Tool, Online Resource.* Available online at: https://sites.uwm.edu/haas/home/rhino_emt/ (accessed January 30, 2020).

Haas, T. C. (2021). *Users Manual.* Available online at: https://sites.uwm.edu/haas/home/idusers/ (accessed May 10, 2021).

Haas, T. C., and Ferreira, S. M. (2018). Finding politically feasible conservation strategies: the case of wildlife trafficking. *Ecol. Appl.* 28, 473–494. doi: 10.1002/eap.1662

Hanneman, R. A., and Shelton, C. R. (2011). Applying modality and equivalence concepts to pattern finding in social process-produced data. *Soc. Netw. Anal. Mining* 1, 59–72. doi: 10.1007/s13278-010-0009-1

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., and Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2018093118. doi: 10.1073/pnas.2018093118

IBM (2021). *Guide to SQL: Tutorial.* IBM Corporation. Available online at: https://www.ibm.com/docs/en/informix-servers/12.10/12.10?topic=programming-guide-sql-tutorial (accessed April 9, 2021).

Institute for Scientific Information (ISI) (2010). *ISI Web of Science.* London: Thompson Reuters.

IUCN/SSC (2007). *Regional Conservation Strategy for the Cheetah and African Wild Dog in Eastern Africa.* Gland: IUCN Species Survival Commission.

Karasneh, Y., Ibrahim, H., Othman, M., and Yaakob, R. (2009). "Integrating schemas of heterogeneous relational databases through schema matching," in *Proceedings of IIWAS2009* (Kuala Lumpur).

Khan, S., and Safyan, M. (2014). Semantic matching in hierarchical ontologies. *J. King Saud Univ. Comput. Inform. Sci.* 26, 247–257. doi: 10.1016/j.jksuci.2014.03.010

Kiasatpour, S., and Whitfield, C. (2008). Politics in action: wolves, snowmobiles, and bison and the greater Yellowstone ecosystem. *Polit. Sci. Polit.* 41, 193–206. doi: 10.1017/S104909650808030X

Kupschus, S., Schratzberger, M., and Righton, D. (2016). Practical implementation of ecosystem monitoring for the ecosystem approach to management. *J. Appl. Ecol.* 53, 1236–1247. doi: 10.1111/1365-2664.12648

Laituri, M. J., Linn, S., Fassnacht, S. R., Venable, N., Jamiyansharav, K., Ulambayar, T., Allegretti, A. M., Reid, R., and Gernandez-Gimenez, M. (2015). "The MOR2 database: building integrated datasets for social-ecological analysis across cultures and disciplines," in *Proceedings of the Trans-disciplinary Research Conference: Building Resilience of Mongolian Rangelands* (Ulaanbaatar). Available online at: https://dspace.library.colostate.edu/bitstream/handle/10217/181708/CONF_MOR2-2015-ENG5-1Laituri_etal.pdf?sequence=1&isAllowed=y (accessed January 30, 2020).

Laurila-Pant, M., Lehikoinen, A., Uusitalo, L., and Venesjärvi, R. (2015). How to value biodiversity in environmental management? *Ecol. Indic.* 55, 1–11. doi: 10.1016/j.ecolind.2015.02.034

Learn, J. R. (2017). *Poaching Isn't the Cheetah's Only Problem.* Smithsonian.com. Available online at: www.smithsonianmag.com/science-nature/poaching-isnt-rare-cheetahs-only-problem-180962808/ (accessed January 30, 2020).

Leenhardt, P., Teneva, L., Kininmonth, S., Darling, E., Cooley, S., and Claudet, J. (2015). Challenges, insights and perspectives associated with using social-ecological science for marine conservation. *Ocean Coast. Manage.* 115, 49–60. doi: 10.1016/j.ocecoaman.2015.04.018

Lima, J. M. T., Valle, D., Moretto, E. M., Pulice, S. M. P., Zuca, N. L., Roquetti, D. R., et al. (2016). A social-ecological database to advance research on infrastructure development impacts in the Brazilian Amazon. *Sci. Data* 3:160071. doi: 10.1038/sdata.2016.71

Ma, X., Pang, H., and Tan, K. (2004). "Finding constrained episodes using minimal occurrences," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)* (Brighton), 471–474.

Mata-Toledo, R. A., and Cushman, P. K. (2000). *Fundamentals of Relational Databases.* Schaum's Outline Series. New York, NY: McGraw-Hill.

Microsoft (2017). *Query and View Designer Tools (Visual Database Tools).* New York, NY: Microsoft Corporation. Available online at: https://docs.microsoft.com/en-us/sql/ssms/visual-db-tools/query-and-view-designer-tools-visual-database-tools?view=sql-server-ver15 (accessed April 6, 2021).

Miyasaka, T., Le, Q. B., Okuro, T., Zhao, X., and Takeuchi, K. (2017). Agent-based modeling of complex social-ecological feedback loops to assess multi-dimensional trade-offs in dryland ecosystem services. *Landsc. Ecol.* 32, 707–727. doi: 10.1007/s10980-017-0495-x

Moe, T. M. (2005). Power and political institutions. *Perspect. Polit.* 3, 215–233. Available online at: https://www.jstor.org/stable/3688027

Nielsen, K., Andersen, T., Jensen, R., Nielsen, J. H. and Chorkendorff, I. (2013). An open-source data storage and visualization back end for experimental data. *J. Lab. Autom.* 19, 183–190. doi: 10.1177/2211068213503824

O'Conner, J. (2006). *Using Java DB in Desktop Applications.* Oracle Technology Network. Available online at: www.oracle.com/technetwork/articles/java/javadb-141163.html (accessed January 30, 2020).

Pankowski, M. (2017). *How to Organize SQL Queries When They Get Long.* LearnSQL. Available online at: https://learnsql.com/blog/5-tips-managing-long-sql-queries/ (accessed April 9, 2021).

ProQuest (2010). *CSA Sociological Abstracts.* Ann Arbor, MI: ProQuest.

Ribary, M. (2020). A relational database of Roman law based on Justinian's digest. *J. Open Human. Data* 6:5. doi: 10.5334/johd.17

Rissman, A. R., and Gillon, S. (2017). Where are ecology and biodiversity in social-ecological systems research? A review of research methods and applied recommendations. *Conserv. Lett.* 10, 86–93. doi: 10.1111/conl.12250

Rosenbaum, P. R. (2002). *Observational Studies, 2nd Edn.* New York, NY: Springer-Verlag.

Shneider, A. M. (2009). Four stages of a scientific discipline; four types of scientist. *Trends Biochem. Sci.* 34, 217–223. doi: 10.1016/j.tibs.2009.02.002

TMAP (2008). *Tanzania Mammal Atlas Project (TMAP), Part of the Tanzania Mammal Conservation Progam Maintained by the Tanzania Wildlife Research Institute.* Arusha. Available online at: www.tanzaniamammals.org (accessed January 30, 2020).

U.S. National Science Foundation (2014). *Earth to Data: Making Sense of Environmental Observations.* News Release 14–135. Available online at: www.nsf.gov/news_summ.jsp?cntn_id=132973 (accessed January 30, 2020).

Virapongse, A., Brooks, S., Metcalf, E. C., Zedalis, M., Gosz, J., Klisky, A., et al. (2016). A social-ecological systems approach for environmental management. *J. Environ. Manage.* 178, 83–91. doi: 10.1016/j.jenvman.2016.02.028

Wesley, D. (2000). Relational database design. *J. Insur. Med.* 32, 63–70.

Wilson, D. E., and Reeder, D.M. (eds.). (2005). *Mammal Species of the World: A Taxonomic and Geographic Reference, 3rd Edn., Vol. 2.* Baltimore, MD: Johns Hopkins University Press.

Xie, Y., Crary, D., Bai, Y., Cui, X., and Zhang, A. (2019). Modeling grassland ecosystem responses to coupled climate and socioeconomic influences in multi-spatial-and-temporal scales. *J. Environ. Inform.* 33, 37–46. doi: 10.3808/jei.201600337

Zhang, L., and Yi, D. (1998). *Working With Subquery in the SQL Procedure.* NESUG98. Available online at: https://www.lexjansen.com/nesug/nesug98/dbas/p005.pdf (accessed April 9, 2021).