
FOREST OBSERVATORY: A RESOURCE OF INTEGRATED WILDLIFE DATA

Naeima Hamed

School of Computer Science and Informatics
Cardiff University, UK
hamednh@cardiff.ac.uk

Omer Rana

School of Computer Science and Informatics
Cardiff University, UK
ranaof@cardiff.ac.uk

Pablo Orozco-terWengel

School of Biosciences
Cardiff University, UK
orozco-terwengel@cardiff.ac.uk

Benoît Goossens

School of Biosciences
Cardiff University, UK
goossensbr@cardiff.ac.uk

Charith Perera

School of Computer Science and Informatics
Cardiff University, UK
pererac@cardiff.ac.uk

October 13, 2022

ABSTRACT

We propose the Forest Observatory, a linked datastore, to represent knowledge from wildlife data. It is a resource that semantically integrates data silos and presents them in a unified manner. This research focuses on the forest of the Lower Kinabatangan Wildlife Sanctuary (LKWS) in Sabah, Malaysian Borneo. In this region, wildlife research activities generate a variety of Internet of Things (IoT) data. However, due to the heterogeneity and isolation of such data (i.e., data created in different formats and stored in separate locations), extracting meaningful information is deemed time-consuming and labour-intensive. One possible solution would be to integrate these data using semantic web technologies. As a result, data entities are transformed into a machine-readable format and can be accessed on a single display. This study created a semantic data model to integrate heterogeneous wildlife data. Our approach developed the Forest Observatory Ontology (FOO) to lay the foundation for the Forest Observatory. FOO modelled the IoT and wildlife concepts, established their relationships, and used these features to link historical datasets. We evaluated FOO's structure and the Forest Observatory using pitfalls scanners and task-based methods. For the latter, a use case was assigned to the Forest Observatory, querying it before and after reasoning. The results demonstrated that our Forest Observatory provides precise and prompt responses to complex questions about wildlife. We hope our research will aid bioscientists and wildlife researchers in maximising the value of their digital data. The Forest Observatory can be expanded to include new data sources, replicated in various wildlife sanctuaries, and adapted to other domains.

Keywords Wildlife data, Internet of Things, Ontology, Linked data, Search and Discovery, SPARQL, Reasoning. **Links:** Forest Observatory Ontology: <https://naeima.github.io/Forest-Observatory-Ontology/> and <https://biportal.bioontology.org/ontologies/FOO>. **Demo:** <https://youtu.be/UZYA70ph5kE>.

1 Introduction

The ongoing habitat transformation in the Malaysian state of Sabah, located in northeast Borneo, poses grave dangers to various animal species. The state of forests may worsen for a number of reasons, including logging, the fragmentation of natural habitats, and the presence of humans. For example, over the past 50 years, Sabah has experienced a loss of nearly 50% of its forest cover. In addition, other threats to its wildlife, such as poaching and illegal trade, have resulted in an average of approximately one human and 13 elephant deaths each year [17, 31]. Compounding these threats, farmers sometimes resort to lethal measures, such as killing or poisoning elephants to protect their oil palm plantations. To achieve the overarching goal of wildlife conservation, scientists and other wildlife researchers have adopted various methods to track environmental changes. To arrive at prompt and well-informed judgments, it is necessary to draw on the findings of several scientific disciplines and access a variety of data sources, achievable through decision-support tools. Such tools can optimise land use, monitor wildlife species, and manage natural resources. For example, unified access to information about vegetation, grazing locations, and plant types can help to understand the foraging behaviour of free-range herbivores.

During the past 15 years, the Danau Girang Field Centre (DGFC) [1], a scientific research facility in the heart of the Lower Kinabatangan Wildlife Sanctuary (LKWS), has collected a variety of data. Collars equipped with GPS placed on elephants [2], as well as trail images and Lidar observations, provide data that are now available. However, these data are siloed and handled independently by the DGFC research team in terms of maintenance, analysis, and storage. As such, it is challenging for decision-makers to collectively access these data and to search for and discover meaningful information quickly.

To this end, we aimed to (i) find the best way to access these data in a unified manner; (ii) allow for a more in-depth analysis of real-world choices by using data from multiple datasets; and (iii) enable data users and software applications to search through multiple data sources while ensuring interoperability (e.g., where the columns in one dataset correspond to columns in others). We acknowledge the existence of standard guides to ensure that everyone creates data with consistent descriptive metadata. However, humans are susceptible to lexical and typographical errors and sometimes define things differently. As a result, we propose the Forest Observatory, a linked datastore to combine different ecological datasets and facilitate the prompt and efficient exchange of information between people and computer systems—i.e., human-readable documentation and machine-readable files based on a unified research identifier (URI). The Forest Observatory enables in-depth analysis of wildlife data.

We addressed the problem of data heterogeneity by using technologies from the semantic web to integrate various data sources in an interoperable way. Allowing complex and granular searches across disparate data sources would be impossible without semantic modelling. By integrating these sources, one could, for example, select all sensor observations, at a particular time, for elephants near a specific area. We started by constructing a sophisticated ontology named the Forest Observatory Ontology (FOO). Although there are many internet of things (IoT) ontologies and several wildlife ontologies, no model, to our knowledge, has connected these two domains. FOO was developed by adopting classes and relationships from well-defined ontologies—namely, the Sensor, Observation, Sample, and Actuator (SOSA) ontology [3] and the BBC wildlife ontology [4]. To create the Forest Observatory, we populated FOO with four distinct datasets (i.e., soil, vegetation, trail camera images, and GPS collars). These heterogeneous datasets came in various file formats, ranging from CSV to JPEG. To semantically model them, we used programmatically convert their CSV files and the images’ metadata to knowledge graphs in Resource Description Framework (RDF) format before loading them into databases (i.e., triple-stores). Moreover, we implanted rules into these databases to infer new events that were not explicitly expressed in the data. For instance, if it was determined that one species in a particular area was in danger of being poached, other animals nearby may also be in danger.

Our proposed linked datastore enables users to query different data types in a format that can be used as input for further analytical operations. To achieve this, we connected it with Python-executable notebooks, which allowed us to script granular SPARQL searches and obtain information from remotely located datasets in a cohesive structure aligned to the needed inputs of further analytic models. We conducted a backdated online search for similar work on Google Scholar and the ACM and Cardiff University libraries. To the best of our knowledge, no published work has developed an ontology or linked data that integrate IoT and wildlife ontologies. Therefore, this study offers the following contributions to the existing literature:

- Forest Observatory Ontology (FOO), an ontology for describing wildlife data generated by sensors. Innovatively, FOO adapted and linked the SOSA and BBC wildlife ontologies.
- Forest Observatory, a linked datastore that integrates four heterogeneous wildlife datasets- referencing FOO
- A Python-based analytical interface (i.e., proof of concept) that visualises and analyses the Forest Observatory.

This paper is structured as follows: [Section 2](#) reviews some relevant related work [Section 3](#) introduces the Forest Observatory ontology (FOO) and explain its life-cycle. [Section 4](#) describes the Forest Observatory and the proposed semantic data integration for four heterogeneous data sources. [Section 5](#) outlines the Forest Observatory’s structure, evaluates it and analyses its results. [Section 6](#) discusses the proposed system and suggests future work. Finally, [Section 7](#) concludes the paper.

2 Related Work

This section reviews some relevant research initiatives in the context of semantic web technologies. We start with defining the term ontology and its past development approaches. Then, we shed some light on linked data usage in ecological and wildlife data modelling.

2.1 Ontology Definition

The term ontology is borrowed from philosophy and defined as the study of existence. To logically relate to artificial intelligence (AI) applications, ontologies can only represent what exists [5]. In the context of AI, computer science and the semantic web, ontologies represent explicit (i.e., the meaning of all concepts must be clearly defined), formal (i.e., data must be encoded in a machine-interpretable language) specification of widely shared concepts and their relations. More specifically, ontologies can be defined as a collection of related terms and their meanings [6]. They are used to express knowledge of a domain (e.g., health, industry, environment and social sensing) to enable machines to communicate using formal language, which is also delivered by a model and has consensus about ontologies. In principle, ontologies are logically well-defined vocabularies that link various data sources and firmly define their connection. They can be built by using classes, relations, and instances. Data entities are represented as graphs with nodes and edges using the Resource Description Framework (RDF). The RDF model transforms a piece of information into a graph that consists of (subject, predicate, object). An example would be (Soil_pH, hasValue, 4.88). Ontologies could be expressed as a tuple of five elements [7], which could be formulated as:

$$Ontology = (C, HC, R, HR, I) \quad (1)$$

where:

- C = (instances of "rdf:Class") and stands for concepts.
- HC = ("rdfs:subClassOf") stands for concept hierarchy.
- R = (instances of "rdf:Property") and stands for relationships between concepts.
- HR = ("rdfs:subPropertyOf") stands for relationship hierarchy.
- I = ("rdf:type") the instantiation of the concepts in a particular domain.

Classes are groups of objects representing the ontology concepts, featured by name-value pairs of attributes. They can relate to other classes, and these relations are another set of name-value attributes whose values are objects of other classes. Classes and their relationship between attributes entail constraints and statements to build formal axioms. Constraints assert the direction of the relationship between classes and their values. As such, statements declare logical expression to assist in eliminating duplication and contradicting information - enhancing the quality of the ontology. An example of a statement can be *the same animal cannot exist in two exact locations simultaneously*. Nevertheless, ontologies can be populated with instances (individuals) of classes with expressed relations among them. For example, an elephant is an instance of an animal. The resulting product is a *knowledge base* that contains terminological knowledge (i.e., knowledge about classes, attributes and formal axioms that form up ontologies) and assertional knowledge (i.e., knowledge about instances (individuals) that populate ontologies). In the remainder of this section, we review ontology development approaches and past linked data initiatives related to wildlife.

2.2 Ontology Development Methodologies

In computer science, building an ontology without a methodology is known as ontology hacking [15]. For this reason, we scoured previous research archives (such as the ACM digital library and Google scholar) in search of an adaptable methodology. We looked for terms such as "ontology methodology", "ontology development methodology", and "ontology building approaches". Several related articles were retrieved, from which we filtered out the most relevant ones. Among the many existing development methodologies, we elected the most recent ones to compare their features-before selecting the best fit for our proposed ontology. The methodologies researched include -the eXtreme Design (XD) methodology [8], a modular, incremental approach that maps a set of competency questions to one or more Ontology

Table 1: Lists and compares the ontology development methodologies. CQs: Competency Questions, NLS: Natural language Statement. Lightweight: Conceptualisation

Ontologies Development Methodologies	Reference	CQs	NLS	Tabular	Integration	lightweight	Formalisation	Implementation	Evaluation	Documentation	Publication	Maintenance
The eXtreme Design (XD)	[8]	*			*	*	*	*	*			
DILIGENT	[9]		*		*	*	*	*	*			
METHONTOLOGY	[10]	*			*	*	*	*	*			
On-To-Knowledge Methodology (OTKM)	[11]	*			*	*	*	*	*			*
Ontology Development 101	[12]	*			*	*	*	*	*			
NeOn Methodology	[13]	*			*	*	*	*	*			
Linked Open Terms (LOD)	[14]	*	*	*	*	*	*	*	*	*	*	*

Design Patterns (ODPs) [16]- prior to integrating them into the ontology under construction. As such, DILIGENT methodology [9] provides a lenient trial and error approach, recommending the order of discussing, evaluating, justifying, and testing in a use case. METHONTOLOGY methodology [17], on the one hand, proposed a waterfall, incremental development approach, focusing on the lightweight ontology version. Although METHONTOLOG provided detailed guidelines for ontologies development life-cycle, it still falls short in generalising the process to fit various domains. On-To-Knowledge Methodology (OTKM) [11], on the other hand, focused on the ontologies' initial setup, enterprise applications and maintenance. Other well-known methodologies are the 'Ontology Development 101' by Noy et al. [12] and NeON by Suárez-Figueroa et al.[13]. Whilst, the former shed light on ontology conceptualisation; the latter organised the development process into nine distinctive scenarios to cater for wider use cases. More methodologies are reviewed in [15] and semantic web tools in [18]. Recently, the so-called Linked Open Terms (LOT) [14] emerged with much to offer. LOT led by example, reflecting on more than two decades of ontological engineering experience. LOT Originates from Neon methodology[19] and encourages borrowing and reusing classes from fully developed and widely shared ontologies. In addition, LOT advises sharing ontologies in accordance with the Linked Data principles [20] so that the research community and software applications can reuse the ontologies. Since our project's intended outcome was linked data rather than the ontology itself, we chose the LOT methodology to develop FOO. As a result, LOT enabled us to publish FOO following best practices and FAIR principles for the semantic web [21]. Another reason is its adaptability to the ontology requirements process (i.e., allowing competency questions, natural statements and tabular data). Table 1 lists and compares the ontologies development methodologies.

2.3 Wildlife Linked Data

Modern technology helped many conservationists make better decisions about wildlife protection (e.g., preventing poaching of endangered species and predicting wildfires). Tracking devices attached to animals, such as GPS, can be used to track their movements and comprehend how their behaviour relates to their environment [22, 23, 24]. Wireless Sensor Networks (WSN) [25, 26] and IoT devices can collect environmental data, such as soil moisture [27, 28], sound [29], and air quality [30]. Further, wildlife forests can be monitored and recorded using motion-activated trails or wildlife cameras [31]. It is also necessary to apply integration techniques capable of handling the heterogeneous nature of wildlife sensors' data (i.e., data collected by different devices in different formats and stored separately). Scientists and wildlife researchers, as such, work together across disciplines to unify, evaluate, and communicate wildlife sensory data with shared meaning for decision support [32, 33]. Remarkably, semantic web technologies have many features to aid wildlife decision support systems [34, 35]. Linking and harmonising heterogeneous data from disparate sources and formally interpreting exchanged information are a few examples of semantic web capabilities. Several past research initiatives applied semantic web technology to model ecological and wildlife data. In particular, Athanasiadis et al. [36] developed a semantic framework for wildlife conservation focusing on the conservation of the large carnivores in the mountain ecosystems of north Greece. Authors employed ontologies for integrating animal tracking data with ecological niche modelling to perform statistical modelling on habitat suitability. In comparison, Our work differs from [36] not only in location but also in the integration process and the output flexibility. Athanasiadis et al. used geostatistical models to output static reports and visualisation. However, our approach integrated Python executable notebook to enable access to diverse data in a structured manner for flexible and interactive analysis operations. In examining the effects of environmental pollution on the ecosystem, Wang et al. [37] used semantic technology and open-linked data

FOO CLASS	IoT Domain								Wildlife Domain				
	SSN	SOSA	SSNX	IoT-LITE	SAREF	OBOE	S3N	IOT-O	ncbitaxon	ecocore	BBC-WO	African-WO	GeoSpecies
Sensor	*	*	*	*				*					
Observation	*	*	*			*		*					
Observable Property	*	*					*	*					
Feature of Interest	*	*			*		*	*					
Results	*	*	*	*			*	*					
ResultsTime	*	*	*			*		*					
TaxonRank											*		*
TaxonName											*		
Person											*		
Image											*		

Table 2: Lists and compares IoT and wildlife ontologies. SSN [41] =Semantic Sensor Network, SOSA [3] = Sensors, Observations, Samples, and Actuators, S3N [48]=Smart Sensor Network, SSxN[49]= , IoT-Lite [42], IOT-o [50], OBOE[44] =Extensible Observation Ontology, SAREF [43] =The Smart Applications REFERENCE, ncbitaxon[51]= , Ecocore[47]= , BBCwo [4]= BBC wildlife Ontology, African-Wo[52]= African Wildlife Ontology, GeoSpecies ontology [45]

to model wildlife observations to support resource decisions concerning fish and wildlife species and their habitats. Their system also stores and discloses provenance data, making it possible to trace the information’s origin. Although there might exist some similarities between our semantic modelling and [37], we found some fundamental differences in multiple areas, including the data transformation method. Wang et al. used a general-purpose tool to manually convert data into RDF models. On the other hand, we convert data pragmatically using a modular python code that can act as an automatic adapter to transform the data and load them into triple data stores. Semantic modelling of trajectories intersects with ecosystems in tracking animal movement with smart devices. Past research by [38, 39] leveraged the semantic inference capacity for effective knowledge discovery and predictive analytics in the trajectory dynamics and data processing for moving objects. Wannous et al. [40], as such, focused on developing a trajectory ontology that integrates features from three different ontology models (i.e., moving objects, marine environment and Spatio-temporal models). Wannous et al.; transformed their data into OWL ontology using an open-source tool (uml2owl). The main difference between our approaches is that the Wannous et al. model built a domain ontology - integrating various ontologies referencing specific use cases. In contrast, our modelling is based on a specific research methodology that allows linked data to populate a foundational ontology and can be applied to various use cases. Furthermore, our methodology provided additional recommended features such as ontology documentation, publication, and maintenance. Many domain-specific ontologies modelled wildlife and sensor data independently. For instance, the Semantic Sensor Network (SSN) ontology [41] describes the entire process of generating sensory observations. Inside SSN, a self-contained ontology named the Sensor, Observation, Sample and Actuation (SOSA) [3]. SOSA can be a standalone ontology for lighter use instead of reusing the SSN ontology. IoT-lite [42] describes the IoT resources on foundational level. SAREF [43], as such, referenced IoT appliances. OBOE [44] focused on modelling the term *observation* and its *measurement*. Concerning exiting wildlife ontologies, we came across [45, 4, 46, 47], each of which covered different wildlife elements depending on the ontology purpose and use case.

3 Ontology Development

This section introduces the Forest Observatory Ontology (FOO). We discuss the ontology development methodology throughout its life-cycle- from gathering requirements to sharing and maintenance.

3.1 Forest Observatory Ontology (FOO)

A novel upper-level ontology that describes wildlife data generated by remote sensing devices. FOO adopts and merges existing ontological resources (e.g., SOSA and BBC wildlife Ontology) from two domains (i.e., IoT and Wildlife). The primary purpose of FOO is not only to allow unified access to heterogeneous wildlife data but to enable standardised data exchange between different computer systems and applications. More specifically, FOO standardises these data entities and formalises their semantics to integrate heterogeneous wildlife datasets from disparate sources. Moreover, it enables structured data mining, machine learning and artificial intelligence applications. For example, a sensor fitted on an animal collects its geo-location observation. Here, the ontology defines the sensor, animal, geo-location and observation- and how they are related in this context. Besides, FOO links different data sources that are semantically modelled and share the same concepts. For instance, data about animals’ locations can be retrieved by querying

these sensors. Nevertheless, FOO supports inferences (i.e., deducing new knowledge from data) via rules injected directly into its database(s). We built on the Linked Open Terms (LOT) methodology [14] to create FOO[53]. The ontology development underwent four iterative phases. These phases comprise the following: (i) Requirements (ii) Implementation (iii) Evaluation, and (iv) Publication. Figure 1 depicts the development process life-cycle.

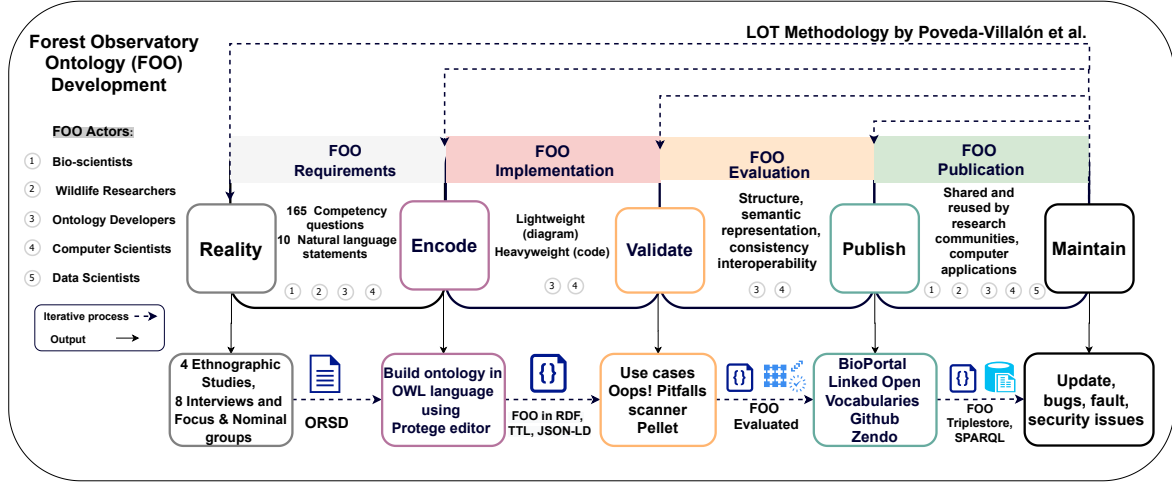


Figure 1: FOO Ontology Development phases, inspired by Linked Open Terms (LOT) methodology [14]

3.2 Ontology Requirements

This phase models the *reality* to generate the ontology requirements specification document (ORSD) [54]. ORSD aggregates the relevant information on FOO’s scope, intended purpose and applied use cases. In this phase, the involvement of domain experts (i.e., bio-scientists and wildlife researchers), ontology developers and computer scientists are crucial to (i) specify the ontology uses cases and (ii) decide the documentations (i.e., datasets) to model. The scope of FOO was defined as linking these two fields, IoT and wildlife.

3.2.1 Functional requirements

Here, we gathered a compiled list of Competency Questions (CQs), Natural Language Statements (NLSs) and various use cases from bio-scientists and wildlife researchers. CQs [55] defined the functional ontology requirements by formulating a set of questions to be answered by the ontology through query languages. On the other hand, NLSs represented short affirmative phrases that aim to define some of the ontology requirements [14]. Nevertheless, the uses cases narrate real-world incidents that the proposed ontology could potentially address. To obtain the requirement mentioned above, we underwent three different activities. The first one involved understanding the wildlife research community through constructing an *ethnography*-informed by casual interviews and observations during data collection. The second activity consisted of semi-structured *interviews* with eight wildlife researchers from Cardiff University in Wales and Danau Girang Field Centre (DGFC) in Sabah, Malaysian Borneo. Thirdly, we conducted *focus and nominal groups*. We prepared three administrative documents for each activity: participant information sheets, consent forms, and demographic forms. The participant’s information sheet briefly described the project’s objective and elaborated on the activity’s procedure. The consent forms allowed us to obtain signed permission to conduct the activities. Lastly, the demographic form gathered non-personal information about participants, including their education level, occupation, and years of experience. Our university research ethics board approved our study requirement collection, and we distributed the documents both digitally and physically. Our target sample size for formal interviews was only six. We were instead able to conduct eight interviews. Similarly, we intended to have at least five individuals in each discussion group (i.e., nominal and focus), and this objective was met. Using Google forms, eleven participant responses were received and analysed. The transcripts of interviews and discussion groups were also revised, and the extracted data was manually coded.

3.2.2 Ethnography

The primary goal of the ethnographic research is to monitor the scientific community’s collection and processing of wildlife data at DGFC [1]. We realised after this experience that our computer science methods (semantic web



Figure 2: Asian elephants herd near DGFC [1]-July 2022



Figure 3: Asian elephants at river

technology) could connect and transform these data into useful information for decision-making. We participated in a variety of activities during our study trip to the DGFC, including:

- **Finding the Sunda pangolin:** Researchers on duty made two trips per day in order to locate a remotely tracked pangolin. They used a VHF antenna to communicate with the scale-mounted receiver on the pangolin. When the antenna approaches the corresponding device, it generates a distinctive sound to signal the pangolin's proximity. Once located, researchers manually record the event's geolocation and local date and time on a paper sheet. At the close of each day, the collected data are loaded into a digital spreadsheet for analysis and storage.
- **Counting wildlife monkeys:** Two groups of researchers took boat trips twice daily in search of two types of wild monkeys, the proboscis and the long-tailed macaques. Once spotted, they count them and list their location and time of the day. This study helped understand the monkeys' behaviours concerning the weather conditions and during different times of the day (i.e., sunrise and sunset). The collected data are then analysed using statistical modelling and visualised images, tables and bar charts.
- **Butterfly trapping:** We joined researchers on a trip to one of the oil palm plantations in the lower Kinabatangan area to hunt butterflies. Researchers have created a catalogue that includes images and information about various types of butterflies. They also carry a net trap and follow a specific rotation to safely catch butterflies. When they catch a butterfly, they visually match it with the catalogue and record its type, as well as the time and date.
- **Finding *Elephas maximus* (Asian elephants):** We went on several boat trips with researchers in search of Asian elephants. A herd was roaming outside the DGFC corridors while the couple was swimming in the lower Kinabatangan river. We took photos, videos, and recorded time and location data.

3.2.3 Interviews

Our eight formal interviewees were all geneticists, biologists, or bio-scientists working in wildlife conservation. Most of the participants we interviewed live in Sabah, Malaysia, except for one scientist who lives in the United Kingdom and visits DGFC regularly. Participants have one to twenty-five years of experience in landscape ecology and conservation biology research. We conducted one-on-one interviews with them at a location of their choice, mostly at their workplace, with one held online. Following our initial interview, we asked the interviewee to recommend the following candidate for an interview. We then sent the interview information sheet and ethics approval information to the nominated candidate two weeks before the scheduled meeting time. Whoever conducted the interview followed the same semi-structured interview schedule, which focused on the types of data they collected and processed and what they hoped to achieve from these data. (i.e., the best way to access the collected data and extract meaningful information to support timely informed decisions). All participants finished their entire interview, and we were allowed to keep their audio recordings for further analysis.



Figure 4: Forest Ontology development, activities collage

Table 3: Functional Requirements Validation Criteria as stated by [55]

Criteria	Results	Bio-scientists	Wildlife Re-searchers	Ontology Devel-opers	Computer Sci-entists
<i>Correctness</i>	All requirements relate to FOO’s concepts.	✓	✓		
<i>Complete</i>	The intended user confirmed FOO’s sufficiency.	✓	✓		✓
<i>Consistent</i>	There were no conflicts between FOO requirements.	✓	✓	✓	
<i>Clear</i>	Each requirement has one precise meaning.	✓	✓	✓	✓
<i>Concise</i>	All requirements were relevant.	✓	✓	✓	✓
<i>Comprehensible</i>	The stakeholders understood FOO requirements.	✓	✓		✓

3.2.4 Focus and Nominal groups

As discussed in the section 4 of this paper, we created a map for Sabah, Malaysian Borneo, as well as various types of data (e.g., elephant GPS collars, soil and vegetation data). We also printed three cards that depict the data and allow participants to make discrete notes and comments on the data. On two consecutive days, we conducted nominal and focus groups. Whilst, the nominal group had six members and one moderator, and the focus group had nine members. They were given a copy of the main map and three cards from each data type. Participants were asked to come up with ideas and potential use cases. Also, to suggest questions that these various data can answer—ideally when linked together, perhaps by location and time. Further, the potential use cases can be applied and how useful they can be in making informed decisions. The focus group provided a list of use cases for FOO and the datasets of interest. We gathered many different perspectives and ideas from the members, some of whom were new to such activities. The focus group yielded a wealth of information, both verbal and written. Participants in the nominal group were allowed to write comments and ideas about the potential uses of the data. Both sessions were approved ethically, consented to, and video recorded. The ORSD was then published in the ontology GitHub repository [53] and was used to evaluate the linked data store.

3.2.5 Functional requirements validation

Signing off the acquired functional requirements requires that they meet the criteria mentioned below, as stated by Grüninger and Fox in [55]. Here, we assessed FOO’s functional requirements before implementation. FOO’s actors verified, to the best of their knowledge, that the requirements are (i) correct, (ii) complete, (iii) consistent, (iv) clear, (v) concise, and (vi) comprehensible. Table 6 has more information regarding the validation criteria.

3.3 Ontology Implementation

Based on the ontology requirements phase results, we determined that the scope of FOO should include IoT and wildlife concepts. For example, our datasets of interest and suggested use cases include data from *sensors* observing *animals* and *land*. The animal GPS collar, for example, tracks an elephant and records geo-location observations for different and equally spaced time intervals and temperature every specified time interval. We needed ontologies that could formalise the following to model domain coverage:

- Different Sensors (i.e., soil, vegetation, trail and GPS collars).
- Wildlife species and their habitats.
- Spatial things.
- Units of measurement.
- Temporal entities

Subsequently, we searched the existing literature in general, and the Linked Open Vocabularies (LOV) [56] in particular for suitable ontologies to reuse (see table 2). After careful consideration, we borrowed classes and relationships from SOSA and the BBC Wildlife Ontology, as they contained sufficient components to model our proposed entities (i.e., instances or individuals). We created and discussed conceptual models (i.e., diagrams) with the ontology stakeholders. Following the graphical representations, we encoded FOO in Web Ontology Language (OWL), edited it with Protégé [57], and wrote pipeline codes in python to serialise the datasets that populate FOO. We publish and maintain all the data and the ontological resources in this [53] GitHub repository.

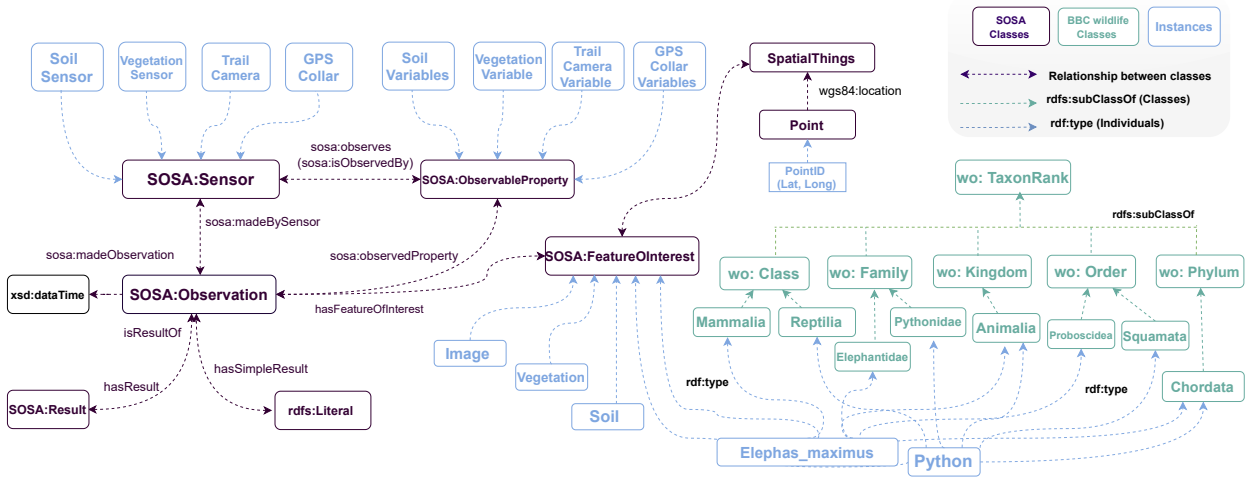


Figure 5: Lightweight version of the Forest Observatory Ontology (FOO), main classes, properties and instances.

Sensor data observations: We adopted classes and properties from SOSA ontology. SOSA is a core ontology originating from the Semantic Sensor Network (SSN) [41] ontology. We acknowledge that there are many ontologies derived from or based on SSN, such as IoT-Lite [42]. However, we adopted the classes and properties directly from SOSA ontology as it was sufficient for the sensors’ observations and data modelling.

Location: We relied on GPS Coordinates as this work deals with outdoor locations.

Temporal aspects: The timestamp is a crucial element in our semantic modelling. It allows us to distinguish between observations in the same dataset and connect them with other observations in different datasets. In this work, we chose to store the timestamp of each observation using XML DateTime data type (i.e., xsd:dateTime). We acknowledge that OWL-Time [58] can be considered to model data and time. However, SPARQL’s Time function is capable of doing the same role. Subsequently, we decided to exclude OWL-Time from FOO.

Units of data: To model our observations unit, we decided to borrow classes from a widely used ontology named Quantities, Units, Dimensions, and Types Ontology (QUDT)[59].

Wildlife features: We inspected multiple wildlife ontologies (see Table2 including the African Wildlife ontology [52] and the BBC wildlife one [4]). Then, we decided to use the BBC ontology for its richness of concepts and properties. (e.g., the hierarchy of Taxonomic rank covers all the animals’ biological classification).

Figure 5 depicts the design of the proposed ontology and Table 5, summarises its content.

3.4 Ontology Evaluation

We started by researching previous ontology evaluation techniques to evaluate our linked data. We comprehended that ontology evaluation focused on checking its quality and correctness throughout its development and after the maturity [60]. We aimed to build the formal ontology in the right way to solve the task. Raad et al. [61] stated four

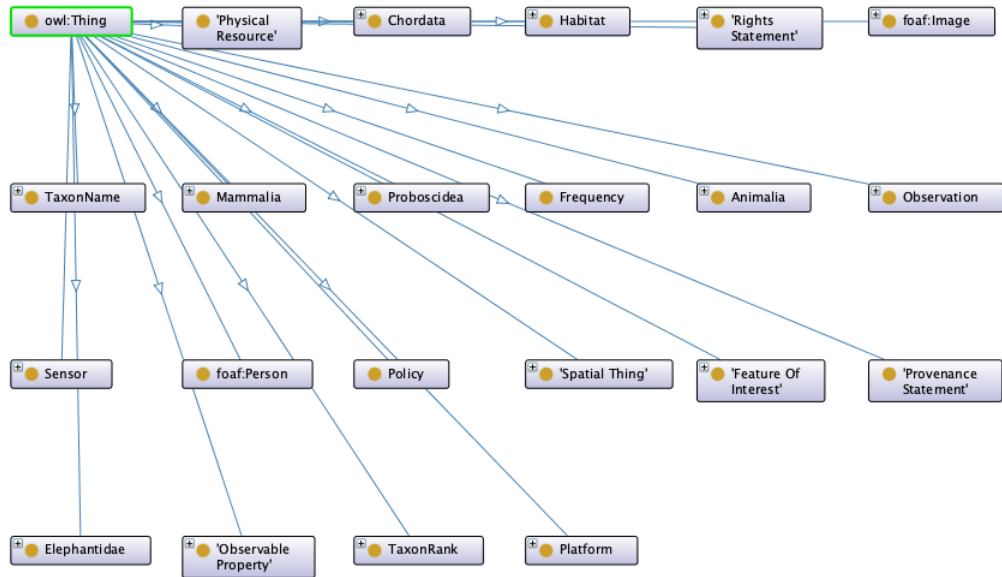


Figure 6: Heavyweight version of the Forest Observatory Ontology (FOO)

ontology evaluations methods from the literature to be (1) gold standards (i.e., comparing it with well-engineered existing ontologies), (2) corpus-based (i.e., checking it against a reference text), (3) criteria-based (i.e., identifying standard qualities and assess them individually) and (4) task-based (i.e., assign it to an application and evaluate its performance). McDaniel et al.[62] described ontology evaluation as a two-way option, the glass-box and the black-box. The former is also called the component evaluation, which examines the ontology incrementally during its life cycle. At the same time, the latter is the task-based approach associated with the performance of a mature ontology in an assigned task or application [63]. Choosing the most suitable approach to evaluate an ontology depends on its intended purpose. Here, our proposed ontology, FOO, aimed to facilitate an application that integrates heterogeneous data sources for decision support. Therefore, we began by evaluating its *structure, semantic representation and interoperability*. To evaluate the structure and semantic representation, the Open-source online scanner, Oops![64] served as the baseline that paved the road for smoother evaluation of the semantic representation- which was achieved using the Delphi method [65, 66]. Following that, the so called black-box, task-based was employed to carry out the linked data store (i.e., FOO instantiated with heterogeneous knowledge graphs) applicability and interoperability (i.e., how well it addresses use cases and how efficient it exchanges data between different computer systems).

3.4.1 Oops! Evaluation

The online freeware scanner [64] detected the structure pitfalls classifying them into three categories, *critical, essential and minor*. FOO had a first rough pass containing different minor pitfalls. Generally speaking, minor pitfalls have no significant impact on ontology performance rather than making it look nice. That said, minor pitfalls such as (i.e., P08 -missing annotation) could degrade the quality of the semantic ontology representation.

P04: the eight minor cases detected the unconnected elements in FOO. We imported directly from the BBC wildlife and the common creative licence website. Although we connected most of the imported elements, few remain for future expansion. Therefore, we take no action to alter block A.

Table 4: Proposed Forest Observatory Ontology: Classes and relationships

OWL Class	URI	Description
<i>Observation</i>	SOSA	<i>Act of carrying out an (Observation) Procedure to estimate or calculate a value of a property of a FeatureOfInterest (e.g., Elephant). Observation can be seen as a placeholder that links relevant information together. As illustrated in Figure 5, in our ontology, observation can be thought of as an ID for each data record. Each row depicts a data record.</i>
<i>ObservableProperty</i>	SOSA	<i>An observable quality (property, characteristic) of a FeatureOfInterest. (e.g., Temperature, humidity, presence)</i>
<i>Sensor</i>	SOSA	<i>Device, agent (including humans), or software (simulation) involved in, or implementing, a Procedure. (e.g., Temperature sensor, humidity sensor, motion sensor). In our model, we have created a unique ID for each sensor based on the platform it is hosted by.</i>
<i>FeatureOfInterest</i>	SOSA	<i>The thing whose property is being estimated or calculated in the course of an Observation to arrive at a Result, or whose property is being manipulated by an Actuator, or which is being sampled or transformed in the act of Sampling. In the context of FOO, Soil is the FeatureOfInterest. Most of the sensors are used to observe a property (phenomenon) of a location (e.g., the moisture of soil).</i>
<i>Result</i>	SOSA	<i>The Result of an Observation, Actuation, or act of Sampling. To store an observation's simple result value one can use the hasSimpleResult property. Result act as a place holder to link related information such as values and units.</i>
<i>SpatialThing</i>	SOSA	<i>A class for representing anything with a spatial extent, i.e., size, shape or position.</i>
<i>TaxonRank</i>	WO	<i>Generic concept for a taxonomic rank such as a Genus or Species.</i>
<i>Class</i>	WO	<i>"associates a taxon rank with a class"</i>
<i>Family</i>	WO	<i>"A family is a scientific grouping of closely related organisms. It has smaller groups, called genera and species, within it. A family can have a lot of members or only a few. Examples of families include the cats (Felidae), the gulls (Laridae) and the grasses (Poaceae)."</i>
<i>Genus</i>	WO	<i>"A genus is a scientific way of showing that species are very closely related to each other. In fact, the first word of the species' scientific name is its genus. So for lions (Panthera leo), Panthera is the genus and tells us that they are closely related to tigers (Panthera tigris), because they share the name"</i>
<i>Kingdom</i>	WO	<i>"Kingdoms are the major categories into which scientists divide up all living things. The main kingdoms are animals, plants, fungi and bacteria, although others exist. Each Kingdom has its suite of defining characteristics - for instance, plants have rigid cell walls, whilst animals do not."</i>
<i>Order</i>	WO	<i>"An order is a scientific way to categorise related organisms. An order is a smaller grouping than a class but bigger than a family or genus. Examples of orders are willows, cockroaches and primates."</i>
<i>Phylum</i>	WO	<i>"A phylum - also known as a division when referring to plants - is a scientific way of grouping together related organisms. All the members of a phylum have a common ancestor and anatomical similarities. For instance, all arthropods have external skeletons. Phyla are large groups and are further subdivided into classes, orders, families and so on."</i>

Table 5: SOSA= <http://www.w3.org/ns/sosa/, WO= <http://purl.org/ontology/wo/

Table 6: FOO's Validation after Implementation

Criteria	Results	Oops! [64]	Pellet [67]	Bio-Scientists
<i>Consistent</i>	There were no conflicts between FOO requirements.	✓	✓	✓
<i>Verifiable</i>	FOO was able to answer complex question.			✓
<i>Clear</i>	Each entity has one precise meaning.	✓		✓
<i>Comprehensible</i>	The stakeholders understood FOO requirements.			✓

P08: reported on missing annotations on thirty-eight cases. We revised the ontology and found that all these elements are duplicates of annotated ones. that is, FOO incurred two copies of them where one copy is missing the annotation. We eliminated duplicates and solved the pitfall case.

P13: flagged another thirty-three cases with undeclared inverse relationships. Given that we adapted these elements, we have no authority in this context to modify them. Therefore, we ignored the P13 result.

P22: We acknowledge that using the same naming conventions in the ontology is one of the W3c recommendations for best practice. However, we anticipated this case due to the nature of FOO, which reuses different ontologies. As we mentioned above at *Block C*, we have no right to modify adopted ontologies' elements.

3.4.2 Pellet Evaluation

Pellet [67] is a open source OWL-DL reasoner. It has a good performance reputation for detecting contradicting facts in ontologies. Pellet can also make inferences and answer SPARQL queries. We used the protege's plug-in pellet version to reason over FOO and check for inconsistencies. Pellet processed FOO in 29 ms, computing the inferences for entities' hierarchies and detecting no contradictions. A screenshot of the reasoner's report is available in FOO GitHub repository [53]. Nonetheless, the inferred axioms have been published as an ontology in Github [53].

3.5 Ontology Publication

Ontologies are typically created in editors and then exported in formats (e.g., Turtle, RDF/XML, JSON-LD) that are challenging for others to use. Researchers sometimes refer to articles or technical reports explaining ontology to solve this challenge. However, papers usually emphasise the scientific contribution rather than defining each

concept in the ontology. Documenting the ontology entities might be a solution. For this reason, semantic web communities have developed tools to assist with ontology documentation. These tools extract annotation properties from OWL ontologies and generate HTML documentation for classes, properties, and instances [68]. We chose to use the WIZARD for DOCUMENTING Ontology (WIDOCO) [69] to document FOO as it builds on Live OWL Documentation Environment (LODE) [70] used in the seven-star linked data model platform [71]. WIDOCO software produces HTML pages containing human and machine-interpretable visualisation and Oops![64] evaluation. Then we made FOO, and its documentation (<https://github.com/Naeima/Forest-Observatory-Ontology/releases/download/v1.0.0/index-en.html>) available in Github [53], not only to maintain it but for also sharing it with the research community, enable them to contribute to its enhancement and interoperability (i.e., use it in software applications) To make FOO findable on the web in different interoperable formats after publishing, we followed W3C best practices and deposited it in the BioPortal repository. FOO [53] is now active online under the creative commons Attribution 4.0 International (CC 4.0). In the following, we describe the datasets of the linked data store and its construction process.

4 Linked Data

This section proposes the linked data approach, depicting the modelling process of the wildlife datasets. We start with describing the data-linking workflows. Then, we introduce the datasets of interest and their semantic data modelling and transformation into knowledge graphs to populate the ontology, resulting in the proposed Forest Observatory.

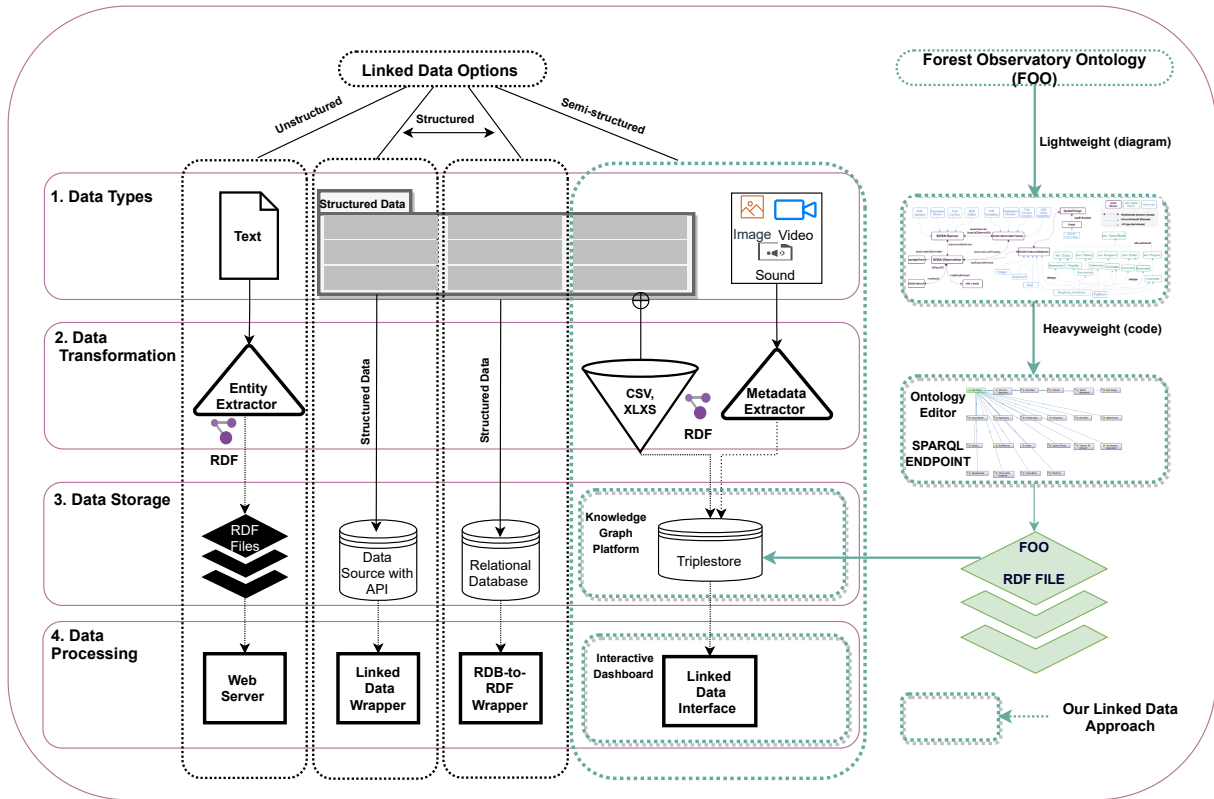


Figure 7: Linked Data options, workflow and our study approach to linking wildlife data.

4.1 Linked Data Approach

The proposed linked data comprises four historical heterogeneous wildlife datasets, stored in disparate locations. Datasets include the soil, vegetation, GPS collar and trail camera images. FOO[53] backed the linking process by defining the datasets’ entities (i.e., concepts, datatype, predicates/properties, instances) and their interrelationships. We only used four datasets in this study due to their availability. Two of the datasets (i.e., soil and vegetation are Open Data from data.gov.uk) whilst the remaining two sensors data sourced from DGFC [1]. Worth mentioning that the Forest Observatory can accommodate more wildlife datasets generated by sensors such as air quality, noise and acoustic sensors. The datasets included in this study were (i) soil data, (ii) Vegetation and site habitats, (iii) GPS collars

Table 7: Soil dataset variables Description

Name	Instance of Class	Data Type	Description
Identifier	sosa:Observation	rdf:type	Unique Sample Identifier
Site	sosa:ObservableProperty	xsd:string	Geographical area/site which samples were taken from
Land_Use	sosa:ObservableProperty	xsd:string	land use of the study plots: Unlogged tropical forest, Logged tropical forest or Oil palm plantation
Plot_Name	sosa:ObservableProperty	xsd:string	Name of the 1 Ha plot sampled
Subplot	sosa:ObservableProperty	xsd:string	Number of the subplot sampled within each 1 Ha plot
Horizon	sosa:ObservableProperty	xsd:string	Soil horizon sampled
Soil_Moisture	sosa:ObservableProperty	xsd:float	Gravimetric soil moisture
Horizon_Depth	sosa:ObservableProperty	xsd:float	Depth of the organic soil horizon sampled
Bulk_Density	sosa:ObservableProperty	xsd:float	Measured Bulk Density of soil sample
Soil_pH	sosa:ObservableProperty	xsd:float	Measured pH of the soil sample
total_C	sosa:ObservableProperty	xsd:float	total carbon content of the soil sample
total_N	sosa:ObservableProperty	xsd:float	total nitrogen content of the soil sample
inorganic_P	sosa:ObservableProperty	xsd:float	inorganic/soluble phosphorus concentration of the soil sample
C:N	sosa:ObservableProperty	xsd:float	carbon to nitrogen ratio of the soil sample
C:P	sosa:ObservableProperty	xsd:float	carbon to inorganic phosphorus ratio of the soil sample

(iv) Images. Creating the Forest Observatory is a non-trivial process due to the dataset’s volume (e.g., big size) and variety (e.g. different formats, CSV, txt, JPG). Besides, assessing and maintaining a good quality knowledge graph is another open challenge. We addressed the data variety issue by transforming the datasets of interest into the Resource Description Framework (RDF) format. We created pipelines (i.e., Python codes) that can also handle significant data volume. All resources are available in our GitHub repository [53]. Another challenge faced was linking the same entities with different names or identifiers. We solved this issue by serialising the datasets with a name or schema that matched those instantiated in the ontology. To illustrate the concept, we populated FOO with the elephant’s GPS collar instances that matched the same data set column headers. Then, we transformed the elephant Abaw’s CSV dataset; the entity was created with the same URI, so that when merged with FOO, it links with the *rdf:type* of *Observation*. Figure13 shows an example of linking an Asian elephant (Abaw) with FOO.

4.2 Soil Data

The data consist of soil characteristics and nutrients for tropical forests in Sabah, Malaysia, both unlogged and logged. Soil properties (ID, Site, LandUse, PlotName, Subplot, Horizon, pH, TotalC, TotalN, TotalP, inorganicP, C-N, Sand, Silt, Clay) extracted from buried ion exchange membranes and soil nutrients (Identifier, Site, LandUse, PlotName, Subplot, NO3N, NH4N, TotalN, Ca, Mg, K, P, Fe, Mn Cu, Zn, B This data is a contribution from the BALI collaboration, which is financed by the UK’s Natural Environment Research Council (NERC)[72]. Modelled datasets in this study can be found at (<https://github.com/Naeima/Forest-Observatory-Ontology/releases/tag/Soil-Data-v1.0.0>).

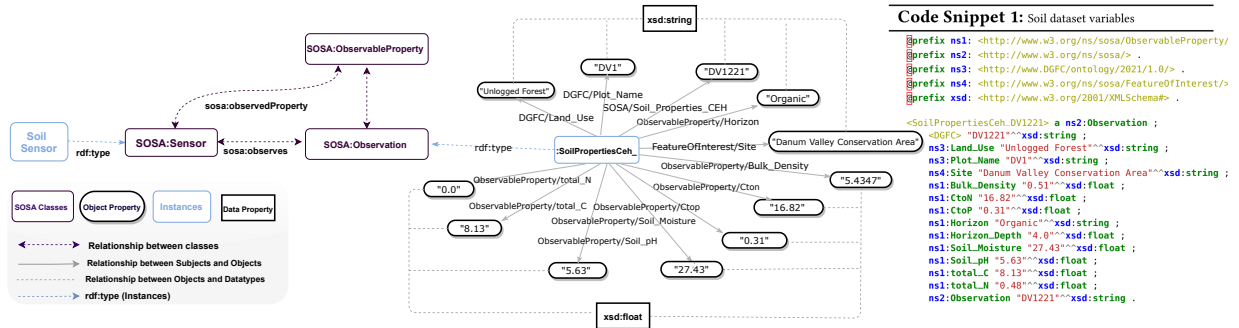


Figure 8: Soil dataset

4.3 Vegetation and Habitat Data

These datasets comprise plant records from forty-nine plots spread among fourteen fragmented forest areas and four continuous forest sites in Sabah, Malaysian Borneo. Live vegetation and deadwood were surveyed at two to three sites at each of the eighteen sites. In addition to vegetation data, the dataset contains measurements of topsoil characteristics, forest structure measures, and metrics of the degree of forest fragmentation in the surrounding landscape of the plots. These data were collected so that studies could be done on (1) the factors that help exotic plant species take over

	Subplot	Soil_Moisture	Horizon_Depth	Bulk_Density	Soil_pH	total_C	total_N	Inorganic_P	C:N	C:P
count	222.00	222.00	222.00	222.00	222.00	222.00	222.00	222.00	222.00	222.00
mean	13.18	26.71	3.74	0.69	5.59	6.04	0.39	33.97	15.04	0.27
std	7.93	9.37	1.91	0.28	0.85	4.61	0.21	51.56	3.52	0.16
min	1.00	7.62	0.20	0.17	3.22	0.83	0.09	3.77	6.65	0.01
25%	6.00	20.56	2.35	0.49	4.92	3.74	0.27	14.34	12.96	0.17
50%	12.00	26.43	3.50	0.66	5.58	4.94	0.34	20.49	14.46	0.24
75%	20.00	31.91	5.00	0.86	6.32	6.33	0.43	32.91	16.47	0.33
max	25.00	65.10	9.50	1.84	7.42	33.45	1.49	571.25	39.59	0.89

Figure 9: Soil data descriptive analysis

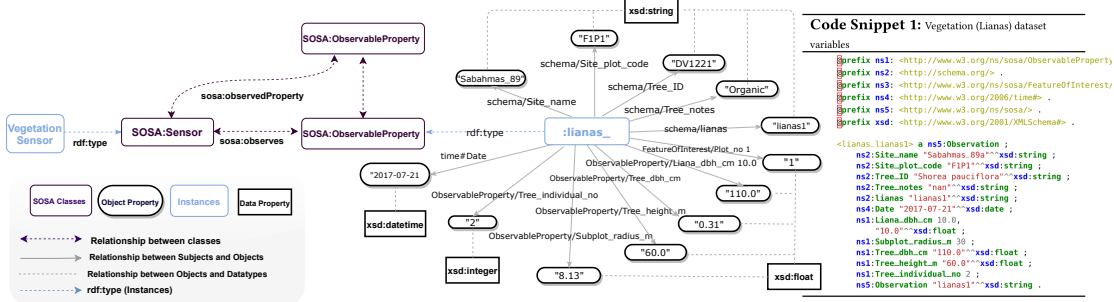


Figure 10: Vegetation dataset variables

fragmented forest areas and (2) the value of conservation set-asides for carbon storage and plant diversity in oil palm plantations.

4.4 GPS Collars

The Danau Girang Field Centre (DGFC) [1] provided us with the collars data for elephants. Collars weighing approximately 14 kg each were officially placed around the elephants’ necks to record numerous metrics (see Table 9) every two hours, including time, geolocation, and temperature [73]. In this study, we modelled twenty-two adult Asian elephants (*Elephas maximus*), fourteen females and eight males; table 9 summarised some GPS telemetry average observations. The collars were created by Africa Wildlife Tracking and fitted ethically by a team of researchers, trackers, and a wildlife vet. Each collar had a Global Positioning System (GPS) receiver and a Very High Frequency (VHF) transmitter. At predetermined intervals, the GPS uploaded each individual’s geographical location information to the Globaltrack server, which could be officially obtained from the Globaltrack website (<http://www.globaltrack.com>). The built-in VHF transmitter aided in retrieving collars that had come off naturally or in locating an individual when additional monitoring was required. The observations in those datasets cover the period range between 2012 to 2018. We refrain from publishing the datasets used in this study because they could be misused to locate an endangered species at risk of poaching.

4.5 Images

A dataset containing one thousand images of *Elephas maxims* was modelled. Before transforming them into RDF graphs, image metadata was extracted and saved as CSV files. The RDF dataset has a unique path that points to the location of the image on a protected cloud server. Figure 14 depicts the dataset entities and the output of the semantic modelling.

5 Forest Observatory Structure and Evaluation

This section examines the creation of the Forest Observatory and its integration with the analytical interface. We then evaluate the Forest Observatory using a task-based methodology. (i.e., assigning a real use case scenario).

Table 8: Vegetation and sites’habitats Data metrics

Name	Instance of Class	Data Type	Description
:DGFC_ID	sosa:Observation	rdf:type	An identifier that is unique for GPS collar sensor observation.
localDate	sosa:ObservableProperty	xsd:date	Local date in Sabah, Malaysia when the GPS collar collects its readings.
localTime	sosa:ObservableProperty	xsd:time	Local time in Sabah, Malaysia when the GPS collar collects its readings.
gMTDate	sosa:ObservableProperty	xsd:date	GMT time in Sabah, Malaysia when the GPS collar collects its readings.
gMTTime	sosa:ObservableProperty	xsd:time	GMT time in Sabah, Malaysia when the GPS collar collects its readings.
lat	wgs84_pos	xsd:float	Latitudinal coordinate of the elephant in question at data collection instant.
long	wgs84_pos	xsd:float	Longitudinal coordinate of the elephant in question at data collection instant.
Temperature	sosa:ObservableProperty	xsd:double	Estimate temperature of the elephant in Celsius at data collection instant.
Speed	sosa:ObservableProperty	xsd:float	Speed of the elephant at current data collection instant.
Direction	sosa:ObservableProperty	xsd:float	Direction that the elephant is travelling at the current data collection instant.
alt	wgs84_pos	xsd:float	Altitude of the elephant in metres at the current data collection instant.
Distance	sosa:ObservableProperty	xsd:float	Distance (m) travelled from the previous data collection instant to the present.
Count	sosa:ObservableProperty	xsd:integer	Observation count per dataset.
HDOP	sosa:ObservableProperty	xsd:integer	Horizontal Dilution of Precision (HDOP) measures GPS accuracy in Latitude and Longitude. The lower value the better precision.

	Plot_no	Tree_individual_no	Tree_dbh_cm	Tree_height_m	Tree_N_lianas	Liana_dbh_cm	Subplot_radius_m	ID1
count	3070.00	3070.00	3070.00	1103.00	3070.00	3070.00	3070.00	3070.00
mean	2.01	1984.24	30.09	20.05	5.74	3.91	25.51	1535.50
std	0.83	1132.74	17.66	10.73	4.65	2.27	4.97	886.38
min	1.00	2.00	10.00	3.00	1.00	2.00	20.00	1.00
25%	1.00	1028.00	17.50	12.00	3.00	2.40	20.00	768.25
50%	2.00	2094.50	26.30	17.00	4.00	3.20	30.00	1535.50
75%	3.00	3022.00	37.00	25.00	7.00	4.60	30.00	2302.75
max	3.00	3895.00	140.00	60.00	31.00	21.80	30.00	3070.00

Figure 11: Lianas data descriptive analysis

5.1 Forest Observatory Structure

Figure 15 illustrates the Forest Observatory’s design. It incorporated FOO and four heterogeneous wildlife datasets. The knowledge graph platform [74] hosted the data and enabled their integration. We used SPARQL [75] to query the linked data via the knowledge graph’s studio that supports many graphical administration tasks. (e.g., data mapping from external sources, virtualisation and interactive visualisation). Forest Observatory Ontology adopted its classes from widely-shared ontologies -including the Sensor Observation, Sample and Actuation (SOSA) [3] and the BBC wildlife (WO) [4] ontologies. The datasets’ entities joined the ontology as individuals. (i.e., instances of relevant classes). These datasets are heterogeneous in context. However, they consider the same geographical location. On the one hand, the soil and vegetation datasets deal with the tropical forests’ properties and nutrients. On the other hand, the animal images and GPS collars datasets focus on tracking the protected animals. This Section discusses each dataset’s elements.

Table 9: GPS Collar Sensor’s metrics

Name	Instance of Class	Data Type	Description
:DGFC_ID	sosa:Observation	rdf:type	An identifier that is unique for GPS collar sensor observation.
localDate	sosa:ObservableProperty	xsd:date	Local date in Sabah, Malaysia when the GPS collar collects its readings.
localTime	sosa:ObservableProperty	xsd:time	Local time in Sabah, Malaysia when the GPS collar collects its readings.
gMTDate	sosa:ObservableProperty	xsd:date	GMT time in Sabah, Malaysia when the GPS collar collects its readings.
gMTTime	sosa:ObservableProperty	xsd:time	GMT time in Sabah, Malaysia when the GPS collar collects its readings.
lat	wgs84_pos	xsd:float	Latitudinal coordinate of the elephant in question at data collection instant.
long	wgs84_pos	xsd:float	Longitudinal coordinate of the elephant in question at data collection instant.
alt	wgs84_pos	xsd:float	Altitude of the elephant in metres at the current data collection instant.
Temperature	sosa:ObservableProperty	xsd:double	Estimate temperature of the elephant in Celsius at data collection instant.
Speed	sosa:ObservableProperty	xsd:float	Speed of the elephant at current data collection instant.
Direction	sosa:ObservableProperty	xsd:float	Direction that the elephant is travelling at the current data collection instant.
Distance	sosa:ObservableProperty	xsd:float	Distance (m) travelled from the previous data collection instant to the present.
Count	sosa:ObservableProperty	xsd:integer	Observation count per dataset.
HDOP	sosa:ObservableProperty	xsd:integer	Horizontal Dilution of Precision (HDOP) measures GPS accuracy in Latitude and Longitude. The lower value the better precision.

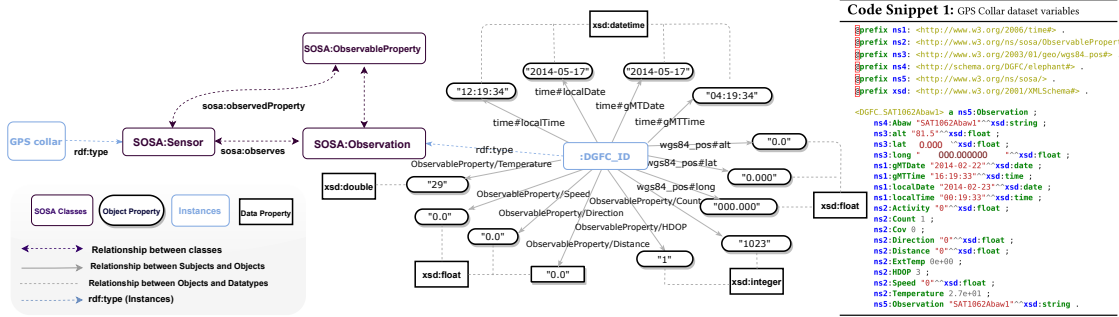


Figure 12: GPS Collar dataset

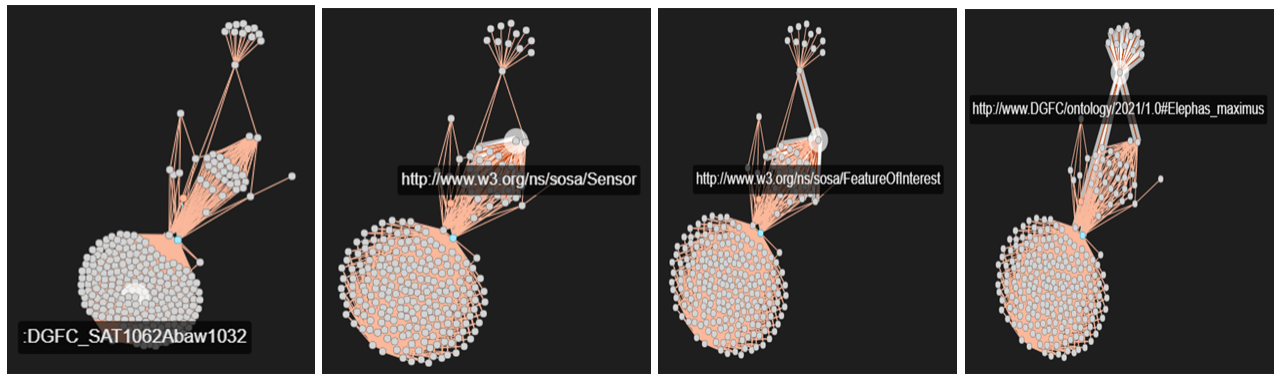


Figure 13: Linking a GPS collar dataset with FOO other entities, from left to right, the first block marks the dataset observations, following that, it links to SOSA sensor through rdf:type Observation. Then the third block connects the feature of interest class to the instance type Elephas maximus which means the Asian elephant

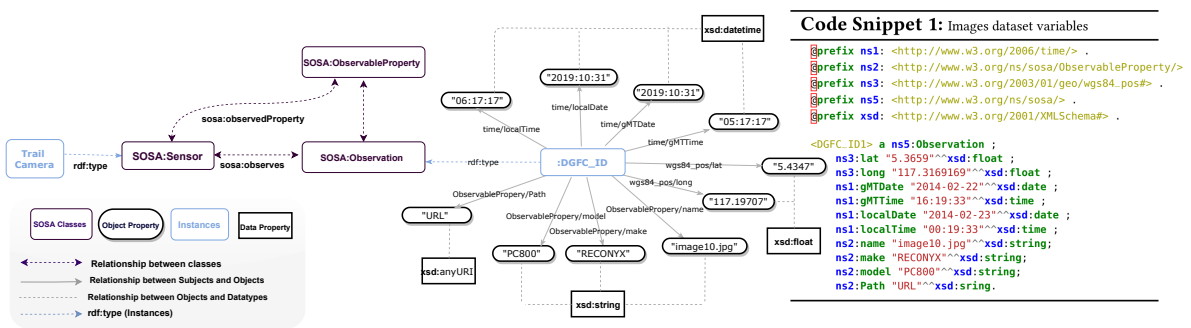


Figure 14: Images dataset

5.2 Forest Observatory Evaluation

To evaluate the Forest Observatory, we sampled a use case 5.3 from the ontology requirements at section 3.2 and evaluated its ability to solve it. Here, we defined multiple competency questions to demonstrate the main contributions of the Forest Observatory. Subsequently, decision-makers can take action based on the evidence obtained.

This use case was designed to verify these contributions, starting by listing three main tasks that the Forest Observatory should fulfil:

- Integrating heterogeneous wildlife data from disparate sources.
- Enabling instant and granular information retrieval.
- Show how reasoning capabilities can infer new facts that were not explicitly expressed in the data.

Table 10: Trail Camera's metrics

Name	Instance of Class	Data Type	Description
DGFC_ID	sosa:Observation	rdf:type	Unique Sample Identifier
Name	DGFC:Name	xsd:string	The named assigned to an image at collection time
Path	sosa:ObservableProperty	xsd:string	The URI to point at the location of the image in secure cloud
localDate	sosa:ObservableProperty	xsd:date	The current local date in Sabah, Malaysia when the GPS collar collects its readings.
localTime	sosa:ObservableProperty	xsd:time	The current local time in Sabah, Malaysia when the image collects its readings.
model	sosa:ObservableProperty	xsd:string	The model of the trail camera used to capture the image
make	sosa:ObservableProperty	xsd:string	The make of the trail camera used to capture the image

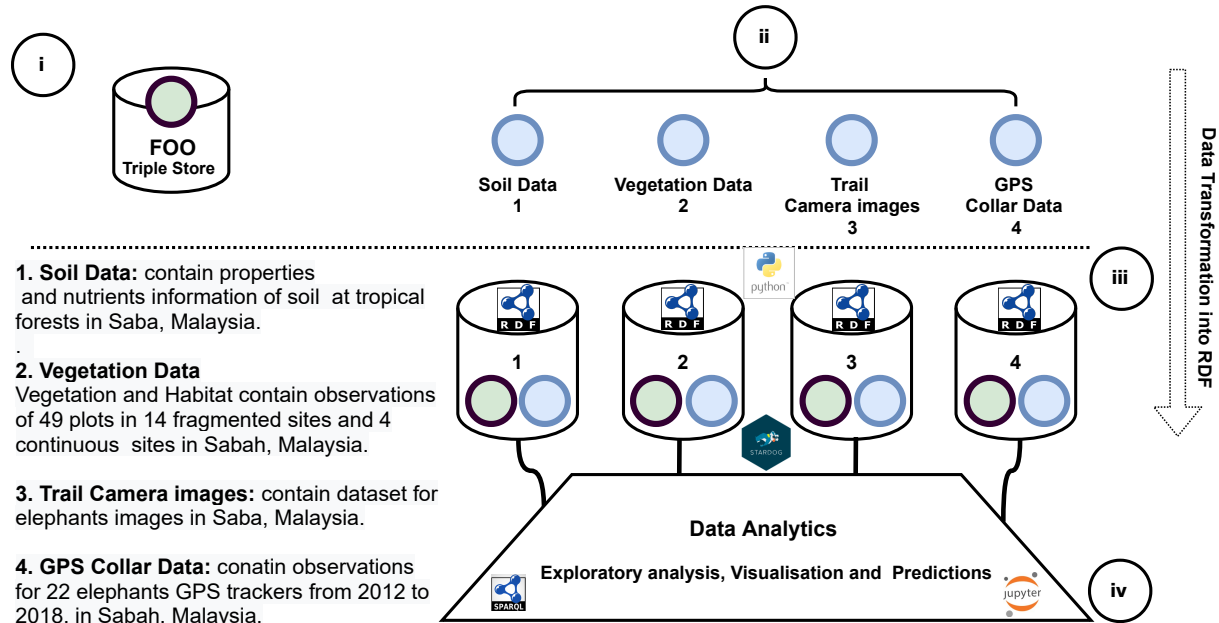


Figure 15: **Forest Observatory Structure:** The system undergoes Four interdependent stages. (1) FOO provides formal definitions for the terms used in the datasets and, thus, describes their relationships with other terms in the ontology. (2) Four raw heterogeneous datasets about soil, vegetation, animal images and GPS collars. We programmatically transformed them from CSV formats into the Resource Description Framework (RDF) model. (3) At this stage, FOO, the ontology, and the semantically enriched data are loaded into triple stores, hosted by a knowledge graph platform [74] and its workspace that enables data query using SPARQL and many more tasks (4) At the final stage, we integrated knowledge graph platform with executable python notebooks for performing data analytics operations of users' choice

To evaluate the Forest Observatory's ability to perform these tasks, we formulated four competency questions from the following use case.

CQ1: What are elephant Jasmin's observations between 2012-02-07 to 2012-02-15? (e.g., Speed, temperature and location)

CQ2: When did elephant Jasmin go near the plantation on 2012-02-07, and how close was it?

CQ3: What is the soil metrics near elephant Jasmin?

CQ4: What are the other elephants near the oil palm plantation?

The first question, Q1 reflects the functionality of integrating the GPS collar information with camera trap images to confirm the elephant's identity and incident urgency. Q2 aimed at justifying elephant behaviour, such as forging and socialising. Question Q3 focuses on the importance of soil conditions near the elephant location to determine the speed and movement behaviour. For example, if the soil was found to be wet, then it is likely that the elephant's legs would be stuck in the mud. Q4 relies on the reasoning capabilities of the semantic web technologies that enable assertive rules to be injected into data. A logical rule may state that if a snare has injured a certain elephant, nearby elephants might be at

risk. These competency questions interrogated our the Forest Observatory through Federated SPARQL queries. Their answers, as such, were retrieved from any available data source integrated with FOO. In other words, separately, we retrieved answers from the relevant knowledge graphs (i.e., FOO populated with elephant Jasmin GPS collar and soil datasets). A sample SPARQL queries are provided below, and the full description, CQs, queries and their answers are accessible on Github [53].

5.3 Rescuing the injured elephant- Use Case

In the Kinabatangan area, many elephants have GPS collars around their neck to locate and monitor their movements. These elephants have distinctive names (e.g., Jasmin, Seri, Sandi, etc.), which are also used to name their GPS collars. Bio-scientists on duty access and visualise the real-time data regularly. Besides, storing historical data for further analysis. One day, while a chief scientist was checking the data, unusual patterns were noticed. Elephant Jasmin's GPS observations were repeating the exact location for two days. Then the scientist sent one of the wildlife officers to check on elephant Jasmin, which was found injured by a snare fitted near an oil palm plantation. The officer immediately informed the manager, who, in turn, contacted the veterinarian to rescue elephant Jasmin.

Proposed Solution: if a particular animal crosses a set of decided geo-location ranges, it should be considered a hazard that requires action.



Figure 16: Use case illustrations- Elephant caught by snare near the oil palm plantation. Injured elephant image was sourced from [76]

Code Snippet 1: What are Jasmin's observations between 2012-02-07 to 2012-02-15?

```

Select *
{ GRAPH <urn:Jasmin> {
?s DateTime:gMTDate ?date;
sosaObservableProperty:Direction ?JasminDirection;
sosaObservableProperty:HDOP ?JasminHDOP;
sosaObservableProperty:Temperature ?JasminTemperature;
pos:alt ?JasminAltitude;
pos:long ?long;
pos:lat ?lat;

FILTER(?date >= "2012-02-07"^^xsd:date
&& ?date <= "2012-02-15"^^xsd:date)}}

```

Code Snippet 2: When did Jasmin go near the plantation on the 2012-02-15 and how close it was?

```

SELECT DISTINCT ?date ?JasminTime ?Distance
{ GRAPH <urn:Jasmin> {?s
DateTime:gMTDate ?date;

DateTime:gMTTime ?JasminTime;
pos:long ?long;
pos:lat ?lat.
?s1 geof:nearby (5.612 117.8436 100 unit:Kilometer).
# Assuming it is the location of the oil palm plantation .
BIND (geof:distance(?s, ?s1, unit:Kilometer) as ?Distance)
FILTER(?date = "2012-02-15"^^xsd:date)}}

```

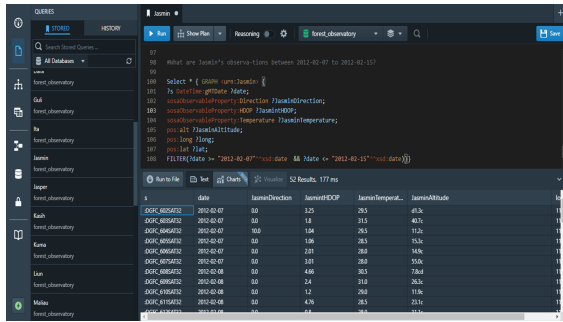


Figure 17: Answer to CQ 1

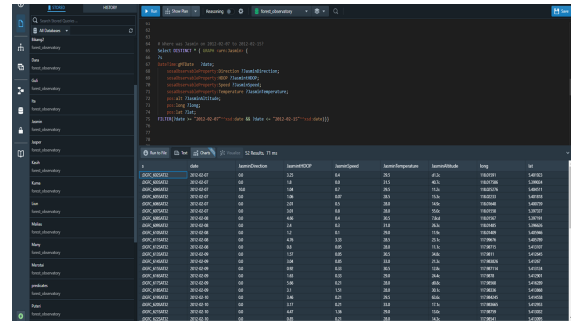


Figure 18: Answer to CQ 2

Code Snippet 3: What are the soil metrics near Jasmin between 2012-02-15-07?

```
SELECT * {?s a sosa:Observation;
schema:Land_Use "Logged Forest"^^xsd:string;
FeatureOfInterest:Site ?Site_Name;
sosaObservableProperty:Bulk_Density ?Bulk_Density;
sosaObservableProperty:CtoN ?CtoN;
sosaObservableProperty:CtoP ?CtoP;
sosaObservableProperty:Horizon_Depth ?Horizon_Depth;
sosaObservableProperty:Soil_Moisture ?Moisture;
sosaObservableProperty:Soil_pH ?pH;
sosaObservableProperty:total_C ?total_C ;
sosaObservableProperty:total_N ?total_N.
```

```
SERVICE <db://FO> {?Jasmin DateTime:MTDate ?date.
FILTER (?date >= xsd:date("2012-02-07")
&& ?date < xsd:date("2012-02-15"))}}
```

Code Snippet 5: Prefixes for CQs SPARQL queries

```
prefix : <http://api.stardog.com/>
prefix stardog: <tag:stardog:api:>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix sosa: <http://www.w3.org/ns/sosa/>

prefix sosaObservableProperty:
<http://www.w3.org/ns/sosa/ObservableProperty/>
prefix pos: <http://www.w3.org/2003/01/geo/wgs84-pos#>
prefix schema: <http://schema.org/>
prefix DGFC: <http://www.DGFC/ontology/2021/1.0/>
prefix DateTime: <http://www.w3.org/2006/time#>

prefix FeatureOfInterest:
<http://www.w3.org/ns/sosa/FeatureOfInterest/>
prefix geo: <http://www.opengis.net/ont/geosparql#>
prefix geof: <http://www.opengis.net/def/function/geosparql/>
prefix unit: <http://qudt.org/vocab/unit#>
```

Code Snippet 4: CQ4: What are the other elephants near the oil palm plantation?

```
SELECT Distinct * {
?ElephantSensor a sosa:Observation;
DateTime:MTDate ?date;
sosaObservableProperty:Direction ?Direction;
sosaObservableProperty:HDOP ?HDOP;
sosaObservableProperty:Speed ?Speed;
sosaObservableProperty:Temperature ?Temperature;
pos:alt ?Altitude;
```

```
geof:nearby (5.612 117.8436 100 unit:Kilometer)
FILTER (?date >= xsd:date("2012-02-07")
&& ?date < xsd:date("2012-02-15"))}}
```

Code Snippet 6: Reasoning (IF-Then-Rule) injected in FOO database

```
prefix rule: <tag:stardog:api:rule:>
[] a rule:SPARQLRule ;

IF {
?s pos:Long ?long;
pos:lat ?lat;
<http://schema.org/DGFC/elephant#Jasmin> ?Jasmin.
?s1
geof:nearby (5.612 117.8436 50 unit:Kilometer).

BIND (geof:distance(?s, ?s1, unit:Kilometer)
as ?Distance).
BIND (?Distance <= 50 as ?Hazard).
FILTER(?Hazard = TRUE)
}

Then {
?s pos:Hazard ?Hazard}
```

5.4 Results

We set up an experiment to evaluate the answers to the use case’s four competency questions. We expected that query results obtained from the Forest Observatory to be *correct, complete and quick*. We ran the queries 50 times (n=50) for each competency question, verified the answer and recorded the response time. Figure 21 plots the response time for each query with the average in milliseconds (ms) of 54.7, 87.29, 259 and 2080 for CQ 1, CQ 2, CQ 3 and CQ 4 respectively. Noticeably, the answers were accurate and correct. CQ 4 has the highest response time due to the amount of requested information per query. Similarly, CQ 3 responses took longer than CQ 1 and CQ 2 for connecting separate databases. Following that, we injected the Forest Observatory database with a reasoning rule shown in code snippet 6. The Rule indicated that if the distance between an elephant and the oil palm plantation is less than 50 Km, then it is a hazard. The correctness and the accuracy of the query results stayed the same whilst the response time was reduced after reasoning 23.

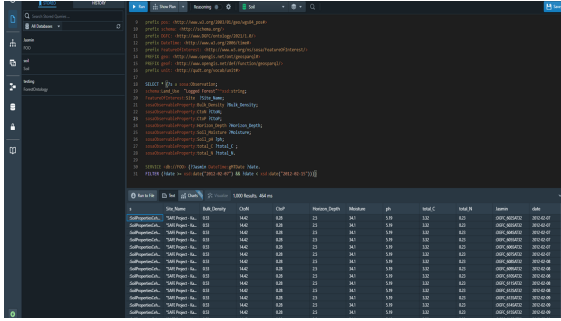


Figure 19: Answer to CQ 3

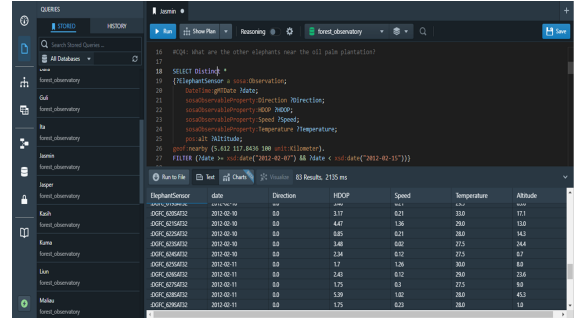


Figure 20: Answer to CQ 4

5.5 Interactive Visualisation

The interactive dashboard connected to the Forest Observatory displays various elephants’ collar GPS data on a map and surveyed sites’ habitats. Visualising data from the linked data store allows end-users to display more data about different elephants (i.e., their location and movements over time). Nevertheless, other data sources such as sites and vegetation could be shown combined as shown in figure 24. The interactive visualisations have entity selection and hover options. The former enables the end-users to select, for instance, the elephant and the other data nearby. Whilst hovering over data points reveals custom information such as the geo-coordinates and timestamps. The use cases of such visualisation could be numerous. To mention a few, (i) combining animal locations and vegetation data (green and water) helps understand animal behaviour during a particular season. (ii) retrieving soil information and site status in a certain area assist in informed actions (e.g. deploying a drone to spray fertilisers and pesticides).

Competency Questions Response Time (ms)

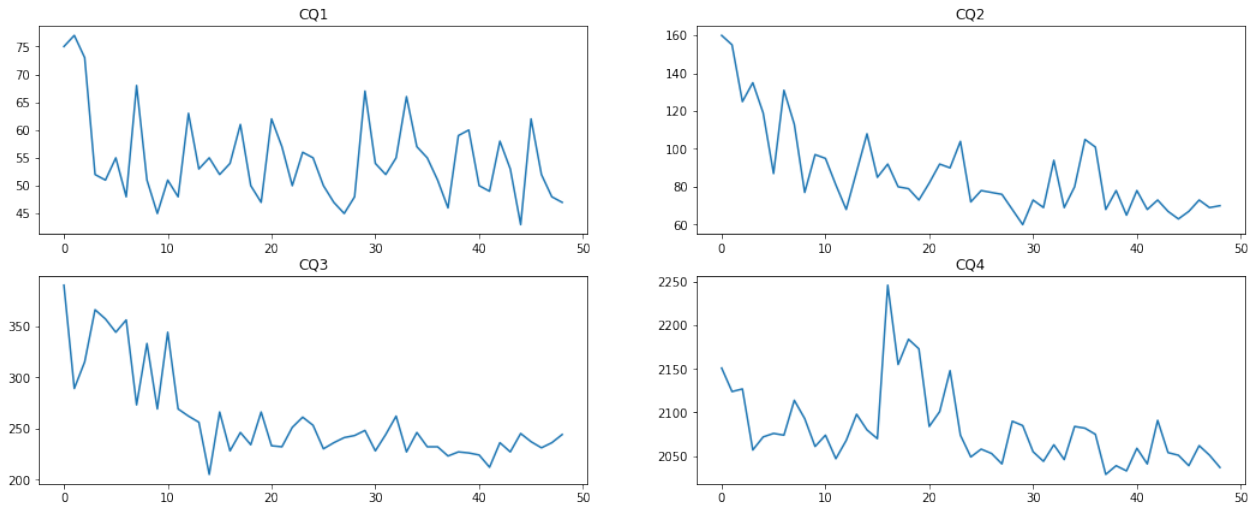


Figure 21: Response time for the four competency questions

6 Discussion

This section interprets the Forest Observatory evaluation results and lessons learnt from the entire experience- acknowledging limitations and recommending future work. The Forest Observatory showed promising results in extracting knowledge from heterogeneous data. In fact, some answers to questions were more valuable when the information was obtained from several sources. Our Forest Observatory performed well in retrieving the use case in section 5.3. Granular data were obtained from disparate sources accurately and reasonably fast. The observations retrieved for the use case 5.3 suggest that incidents may include that the elephant being sick, injured, dead, or its GPS collar being detached. For instance, if the GPS speed is zero, it could mean an error or the device is not sensitive enough (0.001 or 0.01 kw/h), output = 0.0kw/h. Future suggestion compromises that the collar has a heart rate monitor. Nevertheless,

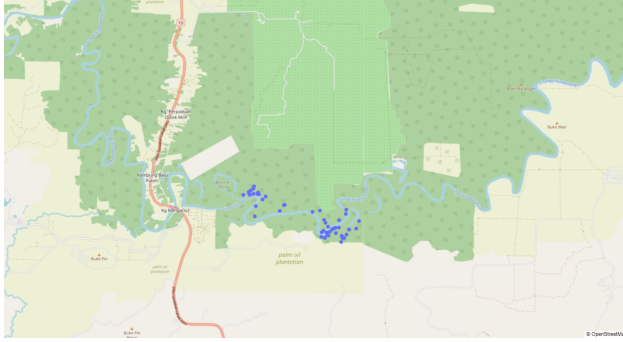


Figure 22: ElephantJasmin movements between 2012-02-07 to 2012-02-15

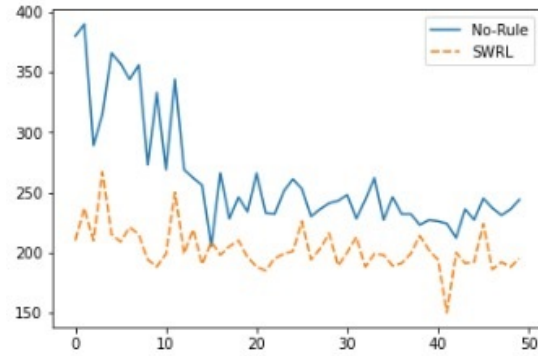


Figure 23: Query response time before and after reasoning

more than 90% of the competency questions used to build the Forest Observatory Ontology (FOO) were validated. That means more use cases can be addressed by the Forest Observatory. One of the lessons learnt from the demonstrated use case 5.3 is that monitoring endangered animals using remote sensing devices could be a practical solution in many ways, such as enhancing knowledge on how elephants utilise the landscape. Nevertheless, the role of the scientists, wildlife officers and veterinarians is equally essential, especially the scientist who possess the skills to read and interpret the GPS tracking data and make the informed decision to rescue the injured elephant. Whilst building the ontology (i.e., FOO) and the linked data store (i.e., Forest Observatory), we can observe that collecting FOO's requirements relied heavily on the collaboration between the bio-scientists and computer scientists. Despite that, obtaining sufficient data to encode the ontology was another challenge. On the one hand, the elephant GPS collar data have sensitive nature. Although we were permitted to model them privately, publishing the entire dataset could potentially put the Asian elephants at risk of poaching. On the other hand, collecting sufficient soil data from DIY sensors was highly challenging. The main reasons were the lack of internet connectivity, intermittent power supply, and the harsh environment (i.e., animals and insects tamper with the devices). Thus, we relied on Open Data to source the soil and vegetation datasets. Worthy mentioning the choice of adopting the LOT [14] methodology for building FOO[53] attributed mainly to the phase of sharing data to the FAIR (Findable, Accessible, Interoperable, Reusable) principles. Such criteria allow the research community to collaborate and contribute to the expansion of FOO. In addition, whilst gathering FOO requirements, we noticed that writing incidents narrated in natural language were equally important as the competency questions. Luckily, the LOT methodology allowed the inclusion of natural language statements during FOO's collaborative requirements phase. Conversely, FOO's implementation process required no interdisciplinary collaboration; instead, it depended on programming and software applications (e.g., Protégé, Python OOPS!, Pellet, and Widoco). Along with its documentation, FOO was shared on GitHub[53], and BioPortal [77]. Once again, publishing our ontology, FOO[53] with LOT methodology promoted ontologies publication, making it more accessible, interoperable, and reusable by broader research communities. That is, adaptable to various research projects and different computer systems. Another evidence of the success of our Forest Observatory is the recent adoption by a research project [78]. This front-ended the Forest Observatory, building AI applications (i.e., chat-bots) to facilitate access to non-domains experts. In brief, the Forest Observatory enabled these heterogeneous ecological datasets to connect meaningfully and allowed reasoning over them to deduce further knowledge.

The limitations of this research are acknowledged. To begin, reusing our proposed ontology [53] may suffer from the heterogeneity of concepts and naming conventions. (i.e., developers may require modifying the ontology entities before interfacing them with their software applications). The Forest Observatory's datasets required security on the production level, especially when data are exchanged between remote SPARQL endpoints. We recommend that future work include adding more ecological datasets and diving deep into building predictive models. Predicting elephants locations could be helpful in many ways. First, to optimise resource management from staff to supplies. Second, to make plans that help protect endangered animals from poaching.

7 CONCLUSION

We created the Forest Observatory, a linked datastore to integrate various wildlife data. It answered complex questions to support decision makers. The Forest Observatory comprised the Forest Observatory Ontology (FOO) instantiated with heterogeneous datasets- modelled as knowledge graphs. The resultant linked datastore contains six million triples and performs multiple tasks. Firstly, end-users can query it directly, as we supplied usage documentation and SPARQL

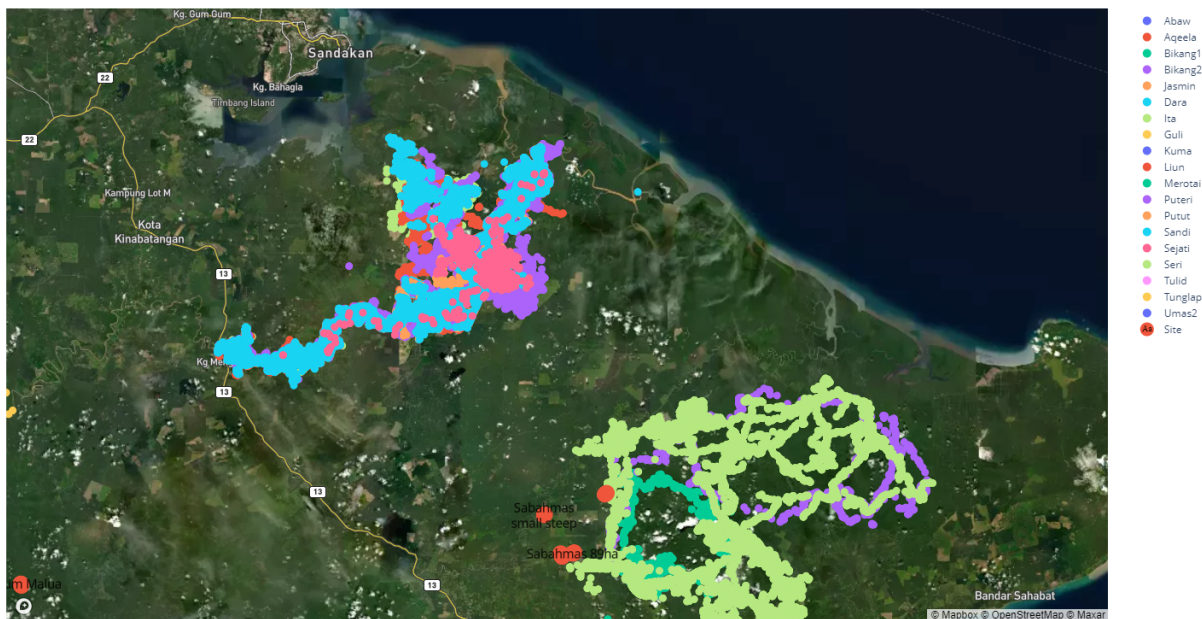


Figure 24: Interactive visualisations for elephants with fitted GPS collars and vegetation sites locations. Demo is available at <https://youtu.be/UZYA70ph5kE>

query examples. Secondly, wildlife researchers can use the reasoning rules injected into the databases to determine which conditions trigger hazards. For example, elephants' locations during a particular time of the year and images of a foreign vehicle nearby could raise suspicion of poaching intent. Finally, domain experts and developers can easily visualise and interact with historical data using the dashboard connected to the Forest Observatory. In the future, we will continue integrating more kinds of wildlife knowledge graphs to build predictive models. Furthermore, the Forest Observatory can be used in other software applications, for example to train artificial intelligence applications such as chatbots and virtual assistants.

References

- [1] Milena Salgado Lynn and Amaziasizamoria Jumail. The danau girang field centre: Field station profile. *ECOTROPICA*, 23(1/2):202103–202103, 2021.
- [2] N. K. Abram, B. Skara, N. Othman, M. Ancrenaz, K. Mengersen, and B. Goossens. Understanding the spatial distribution and hot spots of collared Bornean elephants in a multi-use landscape. *Scientific Reports* 2022 12:1, 12(1):1–16, jul 2022.
- [3] Krzysztof Janowicz, Armin Haller, Simon J.D. Cox, Danh Le Phuoc, and Maxime Lefrançois. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 56:1–10, 2019.
- [4] Naeima/Forest-Observatory-Ontology: The Forest Observatory Ontology (FOO) is an ontology for describing wildlife data generated by sensors., 2022.
- [5] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.
- [6] Yousef Alfaifi. Ontology development methodology: A systematic review and case study. In *2022 2nd International Conference on Computing and Information Technology (ICCIIT)*, pages 446–450. IEEE, 2022.
- [7] Man Tianxing, Mikhail Lushnov, Dmitry I. Ignatov, Yulia Alexandrovna Shichkina, Natalia Alexandrovna Zhukova, and Alexander Ivanovich Vodyaho. An ontology-based approach to the analysis of the acid-base state of patients at operative measures. *PeerJ Computer Science*, 7, 2021.

- [8] Eva Blomqvist, Karl Hammar, and Valentina Presutti. Engineering ontologies with patterns-the extreme design methodology. *Ontology Engineering with Ontology Design Patterns*, (25):23–50, 2016.
- [9] Helena Sofia Pinto, Steffen Staab, and Christoph Tempich. Diligent: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies. In *ECAI*, 2004.
- [10] Mariano Fernández López, Asunción Gómez-Pérez, Juan Pazos Sierra, and Alejandro Pazos Sierra. Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent Systems*, 14(1):37–46, jan 1999.
- [11] York Sure, Steffen Staab, and Rudi Studer. *On-To-Knowledge Methodology (OTKM)*, pages 117–132. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [12] Natalya F Noy, Deborah L McGuinness, et al. Ontology development 101: A guide to creating your first ontology, 2001.
- [13] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.
- [14] María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. Lot: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, 111:104755, 2022.
- [15] Enesi Femi Aminu, Ishaq Oyebisi Oyefolahan, Mohammad Bashir Abdullahi, and Muhammadu Tajudeen Salaudeen. A review on ontology development methodologies for developing ontological knowledge representation systems for various domains. 2020.
- [16] Aldo Gangemi and Valentina Presutti. *Ontology Design Patterns*, pages 221–243. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [17] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo. Methontology: from ontological art towards ontological engineering. 1997.
- [18] Aarti Singh and Poonam Anand. State of art in ontology development tools. *International Journal*, 2(7):96–101, 2013.
- [19] Asunción Gómez-Pérez and Mari Carmen Suárez-Figueroa. Neon methodology for building ontology networks: a scenario-based methodology. 2009.
- [20] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: Principles and state of the art. In *World wide web conference*, volume 1, page 40, 2008.
- [21] Johannes Frey and Sebastian Hellmann. Fair linked data - towards a linked data backbone for users and machines. In *Companion Proceedings of the Web Conference 2021*, WWW '21, page 431–435, New York, NY, USA, 2021. Association for Computing Machinery.
- [22] A. O. Varghese, Arun S. Suryavanshi, and Chandra Shekhar Jha. *Geospatial Applications in Wildlife Conservation and Management*, pages 727–750. Springer International Publishing, Cham, 2022.
- [23] Remo Manuel Frey, Thomas Hardjono, Christian Smith, Keeley Erhardt, and Alex 'Sandy' Pentland. Secure sharing of geospatial wildlife data. In *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, GeoRich '17, New York, NY, USA, 2017. Association for Computing Machinery.
- [24] Gian Pietro Picco, Davide Molteni, Amy L. Murphy, Federico Ossi, Francesca Cagnacci, Michele Corrà, and Sandro Nicoloso. Geo-referenced proximity detection of wildlife with wildscope: Design and characterization. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, IPSN '15, page 238–249, New York, NY, USA, 2015. Association for Computing Machinery.
- [25] Chengjing Ye and Yefen Yang. Wireless sensor network coverage algorithm based on deep forest. In *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture*, AIAM2021, page 1690–1694, New York, NY, USA, 2021. Association for Computing Machinery.
- [26] Vignesh. S, Tarun.M. G, Sankhasubhra Nandi, Sriram. M, and Ashok. P. Forest fire detection and guiding animals to a safe area by using sensor networks and sound. In *2021 4th International Conference on Computing and Communications Technologies (ICCCCT)*, pages 473–476, 2021.
- [27] Devin Bourgeois, Anu G. Bourgeois, and Ashwin Ashok. Demo: Ross: A low-cost portable mobile robot for soil health sensing. In *2022 14th International Conference on COMMunication Systems and NETWORKS (COMSNETS)*, pages 436–437, 2022.

- [28] Daniel Kiv, Garvita Allabadi, Berkay Kaplan, and Robin Kravets. Smol: Sensing soil moisture using lora. In *Proceedings of the 1st ACM Workshop on No Power and Low Power Internet-of-Things*, LP-IoT'21, page 21–27, New York, NY, USA, 2022. Association for Computing Machinery.
- [29] Bethany R Smith, Holly Root-Gutteridge, Hannah Butkiewicz, Angela Dassow, Amy C Fontaine, Andrew Markham, Jessica Owens, Loretta Schindler, Matthew Wijers, and Arik Kershenbaum. Acoustic localisation of wildlife with low-cost equipment: lower sensitivity, but no loss of precision. *Wildlife Research*, 2021.
- [30] Vladimir Dyo, Stephen A. Ellwood, David W. Macdonald, Andrew Markham, Cecilia Mascolo, Bence Pásztor, Niki Trigoni, and Ricklef Wohlers. Wildlife and environmental monitoring using rfid and wsn technology. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys '09, page 371–372, New York, NY, USA, 2009. Association for Computing Machinery.
- [31] K D Cita, R A Adila, R I Hardianto, M F Adib, and L Setyaningsih. Wildlife camera trapping: Estimating the abundance of sumatran tiger's prey in way kambas national park. *IOP Conference Series: Earth and Environmental Science*, 959(1):012020, jan 2022.
- [32] Oscar Novo and Mario Di Francesco. Semantic interoperability in the iot: Extending the web of things architecture. *ACM Trans. Internet Things*, 1(1), mar 2020.
- [33] Jessica L. Oliver, Margot Brereton, David M. Watson, and Paul Roe. Listening to save wildlife: Lessons learnt from use of acoustic technology by a species recovery team. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, page 1335–1348, New York, NY, USA, 2019. Association for Computing Machinery.
- [34] Yves Raimond, Tom Scott, Silver Oliver, Patrick Sinclair, and Michael Smethurst. *Use of Semantic Web technologies on the BBC Web Sites*, pages 263–283. Springer US, Boston, MA, 2010.
- [35] Eva Blomqvist. The use of semantic web technologies for decision support—a survey. *Semantic Web*, 5(3):177–201, 2014.
- [36] Ioannis N Athanasiadis, Ferdinando Villa, Georgina Examiliotou, Yorgos Iliopoulos, and Yorgos Mertzanis. Towards a semantic framework for wildlife modeling. In *EnviroInfo*, pages 287–292, 2014.
- [37] Ping Wang, Linyun Fu, Evan W. Patton, Deborah L. McGuinness, F. Joshua Dein, and R. Sky Bristol. Towards semantically-enabled exploration and analysis of environmental ecosystems. In *2012 IEEE 8th International Conference on E-Science*, pages 1–8, Oct 2012.
- [38] Michael Mireku Kwakye. Semantic data warehouse modelling for trajectories. *arXiv e-prints*, pages arXiv–1904, 2019.
- [39] Georgios M. Santipantakis, George A. Vouros, Christos Doukeridis, Akrivi Vlachou, Gennady Andrienko, Natalia Andrienko, Georg Fuchs, Jose Manuel Cordero Garcia, and Miguel Garcia Martinez. Specification of semantic trajectories supporting data transformations for analytics: The datacron ontology. In *Proceedings of the 13th International Conference on Semantic Systems*, Semantics2017, page 17–24, New York, NY, USA, 2017. Association for Computing Machinery.
- [40] Rouaa Wannous, Jamal Malki, Alain Bouju, and Cécile Vincent. Trajectory ontology inference considering domain and temporal dimensions—application to marine mammals. *Future Generation Computer Systems*, 68:491–499, 2017.
- [41] Semantic Sensor Network Ontology, 2022.
- [42] IoT-Lite Ontology, 2022.
- [43] Laura Daniele, Frank den Hartog, and Jasper Roes. Created in close interaction with the industry: the smart appliances reference (saref) ontology. In *International Workshop Formal Ontologies Meet Industries*, pages 100–112. Springer, 2015.
- [44] Joshua Madin, Shawn Bowers, Mark Schildhauer, Sergeui Krivov, Deana Pennington, and Ferdinando Villa. An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, 2(3):279–296, 2007. Meta-information systems and ontologies. A Special Feature from the 5th International Conference on Ecological Informatics ISEI5, Santa Barbara, CA, Dec. 4–7, 2006.
- [45] GeoSpecies Ontology - Summary | NCBO BioPortal, 2022.
- [46] C Maria Keet. The african wildlife ontology tutorial ontologies. *Journal of Biomedical Semantics*, 11(1):1–11, 2020.
- [47] An ontology of core ecological entities, 2022.
- [48] Samya Sagar, Maxime Lefrançois, Issam Rebaï, Maha Khemaja, Serge Garlatti, Jamel Feki, and Lionel Médini. Modeling smart sensors on top of sosa/ssn and wot td with the semantic smart sensor network (s3n) modular ontology. In *ISWC 2018: 17th Internal Semantic Web Conference*, pages 163–177, 2018.

- [49] Michael Compton, Payam Barnaghi, Luis Bermudez, Raul Garcia-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, et al. The ssn ontology of the w3c semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.
- [50] Nicolas Seydoux, Khalil Drira, Nathalie Hernandez, and Thierry Monteil. Iot-o, a core-domain iot ontology to represent connected devices networks. In *European knowledge acquisition workshop*, pages 561–576. Springer, 2016.
- [51] Yongqun He, Yue Liu, and Bin Zhao. Ogg: a biological ontology for representing genes and genomes in specific organisms. In *ICBO*, pages 13–20. Citeseer, 2014.
- [52] C. Maria Keet. The african wildlife ontology tutorial ontologies: requirements, design, and content. *ArXiv*, abs/1905.09519, 2019.
- [53] Naeima/Forest-Observatory-Ontology: The Forest Observatory Ontology (FOO) is an ontology for describing wildlife data generated by sensors., 2022.
- [54] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. How to write and use the ontology requirements specification document. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 966–982. Springer, 2009.
- [55] Michael Grüninger and Mark S Fox. *Methodology for the design and evaluation of ontologies*. 1995.
- [56] *Linked Open Vocabularies*, 2022.
- [57] *Protégé*, 2022.
- [58] *Time Ontology in OWL*, 2022.
- [59] Jack Spivak and Topquadrant Associate. NASA QUDT Handbook Ontology-based Specification of Quantities, Units, Dimensions and Types The 15th NASA-ESA Workshop on Product Data Exchange Colorado Springs, USA, 21-23 May 2013 Ralph Hodgson TopQuadrant CTO and NASA QUDT Ontologies Lead. 2013.
- [60] Marcus Pendleton, Richard Garcia-Lebron, Jin-Hee Cho, and Shouhuai Xu. A survey on systems security metrics. *ACM computing surveys*, 49(4):1–35, 2017.
- [61] Joe Raad and Christophe Cruz. A survey on ontology evaluation methods. In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2015.
- [62] Melinda McDaniel and Veda Storey. Evaluating domain ontologies: Clarification, classification, and challenges. *ACM computing surveys*, 52(4):1–44, 2019.
- [63] Melinda McDaniel and Veda C Storey. Evaluating domain ontologies: clarification, classification, and challenges. *ACM Computing Surveys (CSUR)*, 52(4):1–44, 2019.
- [64] *OOPS! - OntOlogy Pitfall Scanner!*, 2022.
- [65] Theodore Gordon and Adam Pease. Rt delphi: An efficient,“round-less” almost real time delphi method. *Technological Forecasting and Social Change*, 73(4):321–333, 2006.
- [66] Anusha Indika Walisadeera, Athula Ginige, and Gihan Nilendra Wikramanayake. Ontology evaluation approaches: a case study from agriculture domain. In *International Conference on Computational Science and Its Applications*, pages 318–333. Springer, 2016.
- [67] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Journal of Web Semantics*, 5(2):51–53, 2007.
- [68] G Cota et al. Best practices for implementing fair vocabularies and ontologies on the web. *Applications and practices in ontology design, extraction, and reasoning*, 49:39, 2020.
- [69] Daniel Garijo. Widoco: a wizard for documenting ontologies. In *International Semantic Web Conference*, pages 94–102. Springer, 2017.
- [70] Silvio Peroni, David Shotton, and Fabio Vitali. The live owl documentation environment: a tool for the automatic generation of ontology documentation. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 398–412. Springer, 2012.
- [71] *Linked (Open) Data Finland - Living Laboratory Data Service for the Semantic Web*, 2021.
- [72] Elias et. al. Soil properties across primary forest, logged forest and oil palm plantation in sabah, malaysia, 2018.
- [73] Luke J Evans, Benoit Goossens, Andrew B Davies, Glen Reynolds, and Gregory P Asner. Natural and anthropogenic drivers of bornean elephant movement strategies. *Global Ecology and Conservation*, 22:e00906, 2020.

- [74] The Enterprise Knowledge Graph Platform | Stardog, 2022.
- [75] Dean Allemang. *Semantic web for the working ontologist : effective modeling for linked data, RDFS, and OWL*. ACM books ; 33. Association for Computing Machinery, New York, New York?, third edition edition, 2020 - 2020.
- [76] Raymond Alfred, Laurentius Ambu, SKSS Nathan, and Benoit Goossens. Current status of asian elephants in borneo. *Gajah*, 35:29–35, 2011.
- [77] Forest Observatory Ontology - Summary | NCBO BioPortal, 2022.
- [78] Omar Mussa, Omer Rana, Benoît Goossens, Pablo Orozco-Terwengel, and Charith Perera. ForestQB: An Adaptive Query Builder to Support Wildlife Research. 2022.