FOO: An Upper-Level Ontology for the Forest Observatory

Naeima Hamed [1], Omer Rana [1], Benoît Goossens [2], Pablo Orozco-terWengel [2], and Charith Perera [1]

[1] School of Computer Science and Informatics, Cardiff University, UK
{hamednh, ranaof, pererac}@cardiff.ac.uk
[2] School of Biosciences, Cardiff University, UK
{goossensbr, orozco-terwengelpa}@cardiff.ac.uk

Abstract. Wildlife and preservation research activities in the tropical forest of Sabah, Malaysia, can generate a wide variety of data. However, each research activity manages its data independently. Since these data are disparate, gaining unified access to them remains a challenge. We propose the Forest Observatory Ontology (FOO) as a basis for integrating different datasets. FOO comprises a novel upper-level ontology that integrates wildlife data generated by sensors. We used existing ontological resources from various domains (i.e., wildlife) to model FOO's concepts and establish their relationships. FOO was then populated with multiple semantically modelled datasets. FOO structure and utility are subsequently evaluated using specialised software and task-based methods. The evaluation results demonstrate that FOO can be used to answer complex use-case questions promptly and correctly.

Keywords: Wildlife data · Internet of Things · Ontology · Knowledge Graph · Question-answering
Resource type: Ontology and Knowledge Graph

Ontology's URL https://naeima.github.io/foo_html/
Knowledge Graph's URL https://naeima.github.io/fooKG/
Main website URL: https://www.ontology.forest-observatory.org

1 Introduction

Over the past 15 years, the Danau Griang Field Centre (DGFC)[3], a scientific research facility in Sabah, Malaysia, has collected various data. Collars with GPS chips have been put on elephants, and images from camera traps are also available. However, each research activity maintains its collected data independently, resulting in disparate data. Hence, decision-makers face challenges when accessing these data collectively to search for and discover meaningful information. To address this challenge, we suggest using semantic web technologies, which make it possible to search multiple data sources in a detailed way and to

[3] https://www.dgfc.life/home/

reason about data. Our poster paper contributes an upper-level ontology named the Forest Observatory Ontology (FOO). Following past research methodologies, FOO reuses classes from existing ontologies to connect the Internet of Things (IoT) and wildlife concepts. Then we populated FOO with four semantically modelled datasets to form knowledge graphs. These knowledge graphs enable users to access and query disparate data types in a unified manner, facilitating semantically-enriched information exchange between humans and computer systems.

2 Approach

We searched several previous research archives (e.g., the ACM digital library and Google Scholar) for a suitable methodology. We acknowledge the significance of NeOn Methodology [7]. However, we selected the Linked Open Terms (LOT) methodology by Poveda et al. [6], which builds on Neon Methodology and has features that best match our ontology requirement. For example, competency questions, natural statements and tabular data can all be used at the requirements stage. Figure 1 depicts the development process.

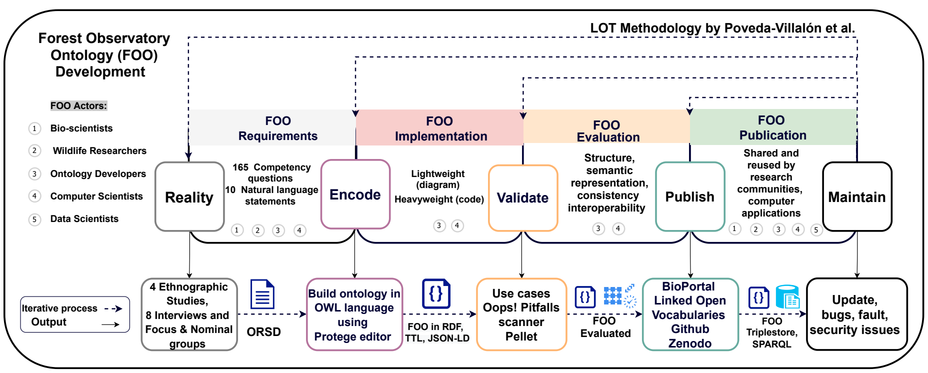

Fig. 1. FOO Ontology Development phases, inspired by Linked Open Terms (LOT) methodology

The Ontology Requirements Specification Document (ORSD) [4] was made to collect information about FOO's scope, its intended purpose, and how it can be used in real life. We compiled 106 competency questions, ten natural language statements (NLS), different use cases from ethnographic studies at DGFC, semi-structured interviews with eight wildlife researchers, and focus and nominal groups at DGFC. For implementation, we searched the existing literature

[4] https://naeima.github.io/FOOBook/lifecycle/requirements.html

for relevant ontologies. We found many of them, such as SAREF[5], IoT-lite[6], SWEET[7] and African wildlife ontology [3]. We chose to reuse the self-contained ontology Sensor, Observation, Sample and Actuator (SOSA) [2] from the second version of the Semantic Sensor Network (SSN) Ontology [1] as it closely matches our requirements. Furthermore, we adopted the BBC Wildlife Ontology[8], which contained sufficient classes to model our wildlife data entities. We discussed the conceptual model (Figure 2) with FOO's actors. Following that, we encoded FOO in the Web Ontology Language (OWL2) (*https://www.w3.org/TR/owl2-overview/*), edited it with Protégé[9], and wrote pipeline codes in Python to map and serialise the datasets that populated FOO. Figure 2 shows FOO's lightweight conceptual model.

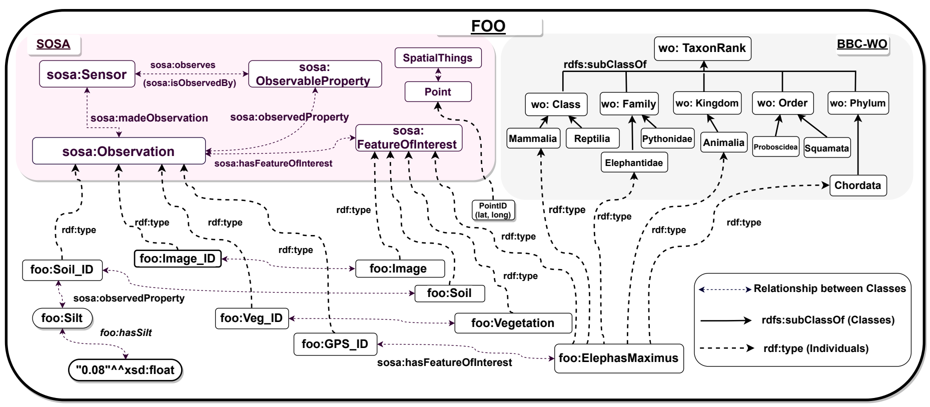

Fig. 2. FOO conceptual model explaining the classes instantiated with data. For example, our soil sensor observation (Soil_ID) is modelled as an instance of the class (sosa:Observation). Then, the observation's metric, (foo:Silt) is modelled as (sosa:ObservableProperty), which has a data property (foo:hasSilt) of type (xsd:float).

We evaluated FOO's structure, semantic representation, and interoperability using open-source online scanners, Oops![10], and Pellet[11] and SPARQL queries that answered competency questions. We instantiated FOO with four distinct wildlife datasets about the forest of Sabah, Malaysia. The datasets compromise sensor observations about (i) soil properties, (ii) GPS elephant tracking collars, (iii) vegetation scanners, and (iv) camera trap images. We programmatically transformed these heterogeneous datasets into RDF graphs. We wrote modu-

[5] https://ontology.tno.nl/saref/
[6] https://www.w3.org/Submission/iot-lite/
[7] https://bioportal.bioontology.org/ontologies/SWEET
[8] https://github.com/rdmpage/bbc-wildlife
[9] https://protege.stanford.edu/
[10] oops.linkeddata.es
[11] github.com/stardog-union/pellet

lar Python codes using the RDFlib[12] library. The approach used is similar to the RDF Mapping Language (RML) mapping technique[13]. The pipeline codes declare the namespaces specific to FOO, iterate through the data, and map the observation id columns to the subjects, the observable property columns to the objects, and their relationship to the predicates. The data source entities were modelled as instances (rdf:type) of FOO's classes, as shown in figure 2. Nevertheless, the mapping codes generate RDF triples that could be serialised into various output formats, such as Turtle, RDF, N3, and JSON-LD. The serialised RDF graphs are loaded into Stardog triple-stores containing FOO to form the knowledge graphs. Each knowledge graph resides in a separate triple-store, representing a data source. Our criteria for creating the federation of knowledge graphs focused on applying a common vocabulary (i.e., a shared ontology) and achieving interoperability (i.e., the disparate knowledge graphs can exchange information using common standards and protocols). To link these knowledge graphs, we used federated SPARQL queries to retrieve data from multiple knowledge graphs simultaneously. The competency question shown in **Listing 1.1** retrieved accurate and prompt information about an Asian elephant's GPS tracking information and the soil condition in a particular area from different knowledge graphs. In order to publish FOO, we used the WIZARD for DOCumenting Ontology (WIDOCO) [5] to generate W3C-compliant documentation. Then, FOO was shared on Github, Bioportal and its dedicated website. FOO's maintenance plan entails routine inspection, scanning, and documenting updates. There will always be issues to resolve, such as bugs or new data to add or remove. Hence, we rely on Github[14] for maintenance, collaboration, and version control. A noteworthy research project by Mussa et al. [4] implemented an AI application, specifically a chatbot, to enhance access to FOO by non-domain experts. As such, we encourage contributions from the research community to support us in extending FOO and identifying additional use cases.

Listing 1.1. What is elephant Aqeela's GPS collar information on 13 November 2011 and the soil sensor information installed at Danum Valley Conservation Area?

```
Prefix foo: <http://www.ontology/ns/foo/1.1#>
Prefix sosa: <http://www.w3.org/ns/sosa/>
Prefix xsd: <http://www.w3.org/2001/XMLSchema#>
Prefix dgfc: <http://www.w3.org/schema.org/dgfc/elephant#>
Prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#>

SELECT DISTINCT * {
  ?UniqueID a  sosa:Observation;
   sosa:madeObservation  dgfc:Aqeela;
   foo:hasSpeed  ?Speed;
   wgs84_pos:lat  ?Lat;
   wgs84_pos:long  ?Long;
   sosa:resultTime  "2011-11-13"^^xsd:date.
```

[12] https://github.com/RDFLib/rdflib
[13] https://rml.io/specs/rml/
[14] https://github.com/Naeima/Forest-Observatory-Ontology

```
foo:ElephasMaximus a ?info;
{SERVICE <username:password@https://[host].stardog.cloud:
port/Soil/query>
{?Soil_ID a sosa:Observation ;
?Site "Danum Valley Conservation Area" ;
foo:hasSilt ?Silt;
foo:hasSoil_pH ?pH.}}
Limit 1
```

3 Conclusion and Future Work

We propose the Forest Observatory Ontology (FOO), an upper-level ontology that semantically integrates heterogeneous wildlife data. It provided answers to complex questions to aid bio-scientists and wildlife researchers in making informed decisions. We instantiated FOO using diverse datasets modelled as RDF graphs. The resultant knowledge graphs contain six million triples capable of performing various operations. First, end-users can remotely query them, as demonstrated in our usage documentation and SPARQL query examples. Secondly, wildlife researchers can incorporate reasoning rules to assert conditions that constitute a threat to wildlife. In the future, we plan to use FOO's knowledge graphs for predictive analytics.

References

1. Haller, A., Janowicz, K., Cox, S.J., Lefrançois, M., Taylor, K., Le Phuoc, D., Lieberman, J., García-Castro, R., Atkinson, R., Stadler, C.: The modular ssn ontology: A joint w3c and ogc standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web **10**(1), 9–32 (2019)
2. Janowicz, K., Haller, A., Cox, S.J., Le Phuoc, D., Lefrançois, M.: SOSA: A lightweight ontology for sensors, observations, samples, and actuators. Journal of Web Semantics **56**, 1–10 (2019)
3. Keet, C.M.: The african wildlife ontology tutorial ontologies. Journal of Biomedical Semantics **11**(1), 1–11 (2020)
4. Mussa, O., Rana, O., Goossens, B., Orozco-terWengel, P., Perera, C.: Forestqb: An adaptive query builder to support wildlife research. arXiv preprint arXiv:2210.02640 (2022)
5. Peroni, S., Shotton, D.: WIDOCO: A wizard for documenting ontologies. In: Proceedings of the 17th International Semantic Web Conference (ISWC 2018). pp. 94–102. ISWC'18, Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00668-3_9
6. Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., García-Castro, R.: Lot: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence **111**, 104755 (2022)
7. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M.: The NeOn Methodology framework: A scenario-based methodology for ontology development. Applied Ontology **10**, 107–145 (2015). https://doi.org/10.3233/AO-150145