Ecolink: Towards a Knowledge Graph Schema for Complex Environmental Systems

Tim Alamenciak[1(✉)], Carlos Alberto Arnillas[2], Harry Caufield[3], Katherine Compton[4], Kian Drew[5], Robert Frühstückl[6], Tina Heger[7], Birgitta König-Ries[8], Chris Mungall[3], Sierra Moxon[3], Justin Reese[3], Jordan Tardif[9], and Lars Vogt[10]

[1] Carleton University, Ottawa, ON K1S 5B6, Canada
TimAlamenciak@cunet.carleton.ca
[2] Department of Physical and Environmental Sciences, University of Toronto at Scarborough, Toronto, ON M1C 1A4, Canada
[3] Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[4] H. G. Thode Library of Science and Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada
[5] School of Environment, Resources and Sustainability, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[6] Johannes Kepler University Linz, Linz, Austria
[7] Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
[8] Faculty of Mathematics and Computer Science, University of Jena, Jena, Germany
[9] University of Waterloo, Waterloo, ON N2L 3G1, Canada
[10] TIB Leibniz Information Centre for Science and Technology University Library, Hannover, Germany
http://timalamenciak.github.io

Abstract. Research findings in ecology have the potential to drive evidence-based actions that could reverse biodiversity decline, inspire nature-based solutions to climate change and enhance restoration of severely degraded waters and lands. However, publishing findings in peer-reviewed papers alone is not sufficient to turn ecological research into action, as evidenced by the burgeoning field of translational ecology. Scholarly literature remains inaccessible to many conservation and restoration practitioners. While the open access publishing movement has increased the availability of research, the knowledge is still poorly indexed and unstructured, leading to inadequate findability.

We present a solution to these challenges in the form of the Ecolink Model (ELM) – an open-source schema for creating knowledge graphs that describe environmental variables, ecological processes and the relationships between them. Drawing on core concepts from ecological modeling and advances in biomedical knowledge synthesis, we outline a model written in LinkML – a domain-agnostic data modeling language – that captures the relationships at the heart of complex systems,

W.-T. Balke et al. (Eds.): TPDL 2025, CCIS 2694, pp. 353–366, 2026.
https://doi.org/10.1007/978-3-032-06136-2_33

thereby providing a structure for knowledge graphs. ELM establishes a consistent and reusable format that enables the discovery of new connections and presents knowledge in an easily searchable, intuitive way. Knowledge graphs that are constructed using ELM have the potential to enable restoration and conservation practitioners to easily access relevant research findings, to unveil new insights using graph data science techniques and drive an AI interface to provide plain-language access to ecological knowledge as described in the graph.

Keywords: ecology · restoration ecology · conservation · knowledge graph · knowledge representation · semantic modeling

1 Introduction

Ecological research can contribute valuable knowledge to restoration ecology and conservation practitioners who are working at the frontline of global ecosystem collapse, but the research is often buried in journals or scattered across multiple, inaccessible sources. Scientific data and knowledge should be findable, accessible, interoperable, and reusable (FAIR) [35], but ecological research is unstructured and poorly indexed, hampering findability and its usefulness. As a result, knowledge synthesis in restoration and conservation ecology is a resource-intensive undertaking.

A graph representation of scientific findings in ecology that utilizes ontologies and semantic data schemata has the potential to facilitate data discovery, enable novel theoretical insights through connections between nodes and enable the application of artificial intelligence to ecology broadly. Findings in ecology emerge from the study of complex systems in which heterogeneity is a core characteristic [1,10]. This heterogeneity frustrates syntheses as the expressed knowledge can be difficult to compare across studies. While some tools exist for documenting scientific findings in the Resource Description Framework (RDF), including nanopublications [9] and the super pattern ontology [3], a schema specific to ecology does not yet exist. In this paper we outline Ecolink Model (ELM) - a schema for describing the findings of ecological research. This schema provides a structure to graph databases that seek to describe the relationships between environmental variables and processes, providing insights about complex environmental systems.

ELM applies existing best practices in biological data to ecology, including the use of ontologies (i.e. machine-interpretable knowledge bases of terms, links between them and definitions), structured templates and semantic data schemata (i.e. specifications for the structure of datasets) [8]. These practices are implemented in biology through the Biolink Model - an open-source schema that formalizes the associations between types of biological entities (e.g. genes and diseases; phenotypes and diseases; etc.) [34]. The Biolink Model characterizes relationships between entities in a nested structure. The Biolink Model emerged from an interoperability challenge in biomedicine: many separate labs were producing knowledge graphs, but there was no central structure allowing them to be combined, potentially missing out on insights from synthesis [34]. Much like in ecology, the research findings and data on which they were based were heterogeneous, frustrating synthesis. The Biolink Model has allowed researchers to produce multiple knowledge graphs with diverse focuses such as the synthesis of Covid-19 research [27] and phenotype-genotype relationships to support precision medicine and disease modeling [24]. The model has also led to the creation of a central hub for the construction and use of such knowledge graphs [6].

ELM applies the association form of the Biolink Model to the relationships among environmental variables and environmental processes. By building on existing technologies, we are able to integrate ELM with a software stack that includes a flexible knowledge graph depiction format[1], ontology-driven text extraction[2], and programmatic access to the schema along with validation using LinkML [23][3]. ELM taps into a network of scientists, practitioners and data curators already using the Biolink Model software stack, creating interdisciplinary linkages that will help broadly advance knowledge representation.

ELM can be used to specify graph data models, which allow for the representation of knowledge and empirical data using triples – subject, predicate, object [30]. ELM is most suited for a labeled property graph-style database (e.g. Neo4j). ELM is also paired with its own project ontology, Ecolink Model Ontology (ELMO), which imports terms from well-developed domain ontologies like the environment ontology (ENVO), an ontological representation of NCBI's taxon database (NCBITaxon) and the relation ontology (RO). A significant portion of the work of constructing ELM has involved updating ontologies that may be used to describe ecological processes, such as adding ecosystem management processes to ENVO, as well as gathering terminology for environmental variables and incorporating the IUCN's ecosystem functional group typology into ENVO [16]. While the primary subject of this paper is ELM itself, we will briefly discuss ELMO as well.

1.1 Use-Cases

The quality of a schema is determined by how effectively it can meet the needs of its users. ELM is primarily a tool for modeling findings – it does not attempt to model reality, but rather the findings as depicted in journal articles. The literature we are aiming to capture using the ELM schema focuses on observational or experimental studies in real ecosystems, or on greenhouse studies in which some ecosystem condition is simulated. While ELM can be applied broadly in ecology, to illustrate its usefulness we focus on restoration ecology, for which we describe three possible use-cases:

1. **ELM as building blocks for a knowledge organization system.** ELM and the ontologies it uses can enable restoration practitioners and researchers

[1] https://github.com/biolink/kgx.
[2] https://github.com/monarch-initiative/ontogpt.
[3] https://github.com/linkml/linkml.

to easily search for findings in similar contexts to their own that are relevant to taxa, techniques and processes they are interested in.
2. **ELM as graph data schema.** ELM enables the use of graph data science techniques to discover new linkages between and among environmental variables and processes.
3. **ELM for retrieval-augmented generation.** We intend to power an LLM interface to provide plain-language access to ecological knowledge as described by a graph based on ELM.

1.2 Assumptions and Constraints

ELM is best understood as a schema that can provide the basis for useful interfaces, such as a knowledge organization system [22]. Its goal is to depict the research on complex environmental systems in such a way as to facilitate knowledge discovery. It seeks to model knowledge in academic articles, not to reproduce it entirely in a new form or model the world. Therefore, some assumptions are required to scope the schema.

We draw inspiration from the practice of ecological modeling - creating representations of ecological processes using state variables (i.e. environmental variables) and flows [15]. Such models are very useful as they are able to answer questions about the anticipated behaviour of complex systems. However, they must necessarily include assumptions in order to allow for the depiction of these systems. Such assumptions limit the scope of what can be modeled, but increase the explanatory power within the scope covered [13].

ELM focuses on links among environmental processes and environmental variables [25]. In an ecological model, links between environmental variables (called “state variables”) take the form of mathematical equations that either limit or enhance flows. Broadly speaking, the structure of ecological process associations in ELM take basic flow rate equations from ecological modeling as inspiration for the structure. Specifically, ELM allows for the expression of a second-order flow rate, where the flow rate between two environmental variables is conditioned by up to two other environmental variables C_1 and C_2:

$$rate = \frac{dC}{dt} = k \times C_1 \times C_2$$

In this basic equation, dC refers to the degree of change at each time step dt. That rate is said to be equal to k which can be interpreted as the coefficient of correlation between two environmental variables in a given model. In modeling, these parameters represent environmental variables that have some impact on the direct flow expressed by k. Palmeri et al. [25] write that this expression, and modified versions of it, are at the core of many ecological models. In ELM, the core association (i.e. subject-predicate-object) corresponds to the flow rate k and the slot `conditioned by qualifier` corresponds to C_1 and C_2. However, the correspondences are conceptual only and contain no numerical values.

ELM is a schema that models research findings as published rather than the ecosystems in which those findings occur – an important distinction. A complete ELM entry cannot be directly translated into equations, or mathematically verified. It can, however, provide vital evidence for the construction of ecological models by linking findings across contexts. For instance, a modeler working in the wetland ecosystem and modeling *Phragmites australis* could draw all entries with *Phragmites australis* as the subject or object of an association documented in that ecosystem. As a semantic artifact, ELM's expressiveness is somewhere between metadata (i.e. descriptions of the article including keywords, title, author, etc.) and the full text in terms of level of complexity.

2 ELM Specification and Core Example

ELM is specified in LinkML and submitted on GitHub for integration with the Biolink Model. ELM inherits classes from Biolink, specifically the core association class and qualifiers (see Fig. 1). Qualifier slots combine with other slots to form compound statements. For example, the `direction qualifier` slot combines with the `subject value` slot to form a subject qualified with the direction of change. This allows ELM to specify the nature and direction of the association.

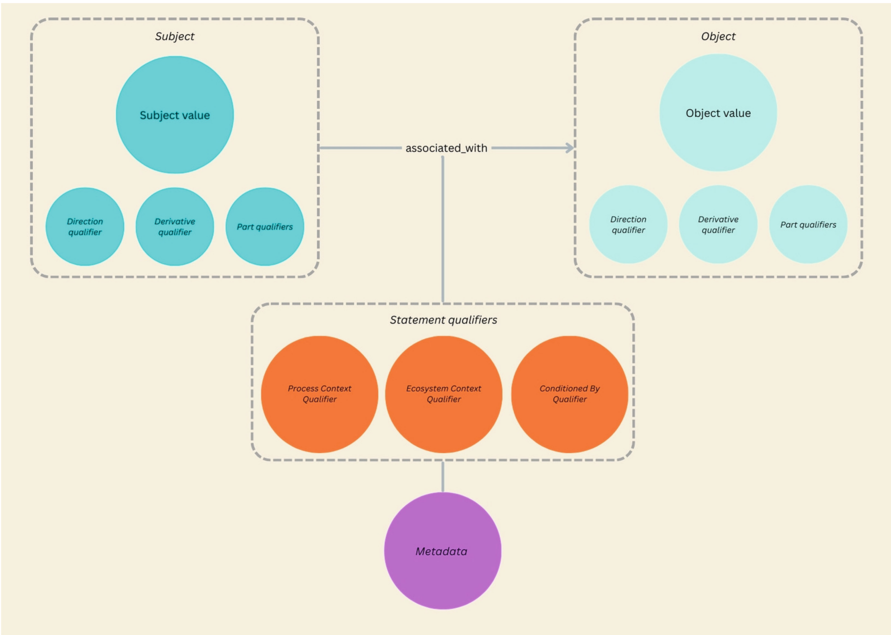

Fig. 1. The overall structure of ELM's association schema. Each circle represents a "slot" in LinkML. This representation shows all possible slots in the schema. Note that the statement qualifiers and metadata apply to the entire association and not simply the edge. It is represented here as attached to the edge because a Neo4j database would model it as such.

Other qualifiers, such as statement qualifiers, apply to the entire statement, providing context in which the statement holds true. In short, qualifier slots allow for the combination of individual ontology terms, increasing the expressivity of ELM. We have placed additional constraints around certain slots in ELM to constrain the possible entries to ones rooted in an ontology. For instance, any values in the `ecosystem context qualifier` slot must have a value that is a subclass of *ecosystem functional group* in the ELM ontology in order for validation to be successful. This combination of a defined graph pattern with slots that restrict the value to instances of specific ontology classes ensures that all data that is modeled using this schema will be semantically interoperable, which facilitates data integration, findability, and reusability. However, this also increases the importance of ensuring that depictions of ecosystem types in the ontology is consistent with current best practices. We will break down the individual slots in the coming sections with reference to an ongoing example, outlined below.

2.1 Core Example: Bare Soil Cover and *Myrmica Scabrinodis* Population

We provide an example that uses ELM to depict one part of the findings from a paper that investigated the effect of sod translocation on the population of *Myrmica scabrinodis*, a species of ant that was the target of a restoration intervention [29]. The paper itself uses a generalized linear mixed model to analyze the relationship between the responses of several environmental variables to a restoration intervention. Bare soil cover is found to have a statistically significant, inverse relationship with the population of *Myrmica scabrinodis*.

This finding means that as bare soil increases, the population of *Myrmica scabrinodis* tends to decrease. The authors suggest this is because *Myrmica scabrinodis* needs the shade provided by plants to survive [29], but for our purposes the relationship and its attribution are sufficient. You can see a representation of the Ecolink entry for [29] in Fig. 2.

The graphical representation shows how the layers of meaning are affected by the structure of ELM. A completed representation of [29] would contain several iterations on the association depicted in Fig. 2 describing the relationships among other variables in the study. There is a representation of the core example in YAML contained in Fig. 3.

2.2 Environmental Variables and Environmental Processes

ELM specifies that each association's subject and object should consist of environmental variables and/or environmental processes. Environmental variables represent some aspect of the state or condition of an ecosystem at a given point in time [15]. Examples of environmental variables include bare soil cover, the population abundance of a particular taxon, or the phylogenetic distance between two taxa. Some environmental variables may be represented using a term from an ontology, while others may have to be post-composed in ELM relating multiple ontology terms to each other. For instance, an ontology could specify a term

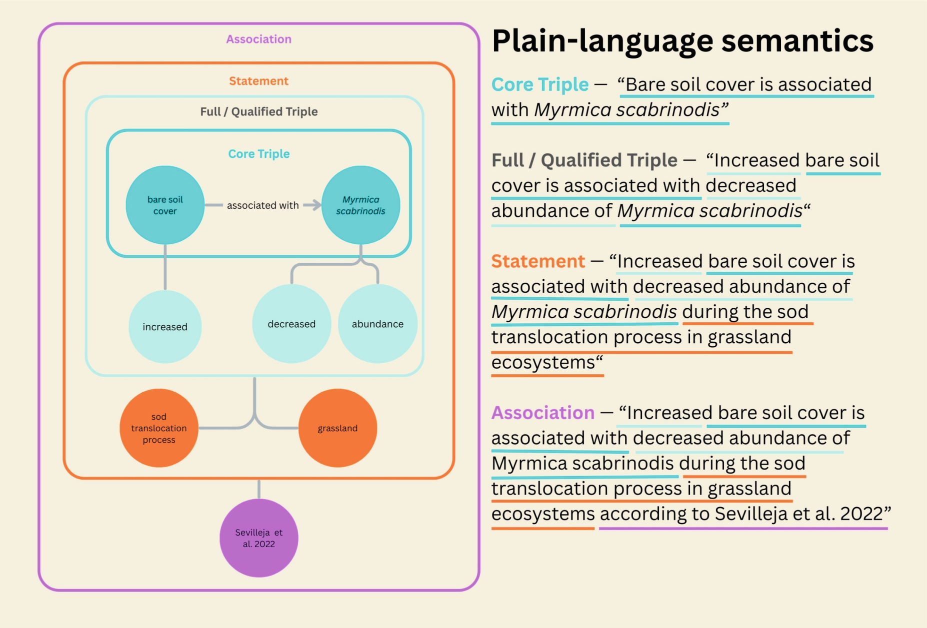

Fig. 2. Biolink Model is based on a nested structure with a core triple in the centre. The subject and object in the triple can be qualified. The whole statement can also receive qualifiers. Everything is wrapped in metadata that provides provenance for the statement. This graphical representation shows the values that would be included in a single ELM entry that is representing one association discovered in Sevilleja et al. 2022.

for *bare soil cover* as "the amount of bare soil in a given area," in which case that term could be used in the subject or object. *Species abundance*, however, would have to be post-composed in ELM as referring to the abundance of a specific taxon by using the `derivative qualifier` slot (Fig. 1). Some terms may be relative and require additional information to have meaning. For instance, *phylogenetic distance* is a measure that only has meaning when that distance is between two or more taxa. In that case, those taxa could be included in the `part qualifier` slots.

Qualifiers combine with the core subject or object and become attached to that element in the graph database. For example, the direction qualifier (Fig. 1) indicates the direction of change that the variable experiences as part of the association (i.e. increase or decrease), as is the case with increased *bare soil cover* in the core example.

The `derivative qualifier` signifies that the full subject should be interpreted as "*qualifier* of *subject*" – for instance, "abundance of *Myrmica scabrinodis*". This enables rich semantic searching, where users could search for the taxon they are interested in, or the qualifier of some taxa, or a combination of both.

Processes are represented in ELM as a series of associations between state variables, where the process in which that association occurs is specified through the `process context qualifier` slot. One process (e.g. *sod translocation process*) can involve numerous associations. These elemental associations may or may not be linked to one another. Processes may be specified as subject or object in the association's core triple (see Fig. 2), or as the `process context qualifier`. That qualifier is used when one process acts as a sub-process of a larger one. For instance, crushing tile drain may be a sub-process of some specific wetland restoration process, in which case crushing tile drain would be either the subject or object and wetland restoration process would be the `process context qualifier`.

2.3 Edge Type: Associated With

ELM uses the `associated_with` edge as defined in the Biolink Model as the sole edge type. This is defined as a relationship between two things, typically statistical in nature, that is weaker than correlation but stronger than relation [34].

The use-cases described in Sect. 1.1 point to a need to gather a wide variety of literature, rather than to deeply and thoroughly model each study. There is a trade-off inherent in this practice: a more detailed graph will take more resources to assemble. By simplifying the edge type in this association, the syntax of the model language is clear and replicable across numerous studies. A more complex edge type that accounted for coefficients may be beneficial, but would also require significantly more time in which to extract information.

2.4 Context and Conditioning Qualifiers

ELM provides three slots for qualifiers that apply to the entire association: `process context qualifier`, `ecosystem context qualifier` and `conditioned-by qualifiers` (Fig. 1). The two context qualifiers describe the conditions in which the association was observed or studied.

An environmental variable or a process can condition the association, which indicates it has some meaningful impact on the association. For instance, in a predator-prey relationship, the association may be conditioned by the population of the predator. This property mimics the concept of "forcing functions" from environmental modeling [15] in that the `conditioned-by qualifier` can be said to have either a dampening or enhancing effect on the association. The direction of that effect is not specified in ELM.

2.5 Ecolink Model Ontology (ELMO)

The Ecolink Model Ontology (ELMO[4]) is a project ontology that imports terms from environment domain ontologies in three main areas: environmental vari-

[4] https://github.com/timalamenciak/elmo.

ables, ecosystem types and environmental processes. Its primary and sole use-case is to power the Ecolink Model. We created the ontology repository in compliance with best practices for ontologies established by OBO Foundry[5], using the Ontology Development Kit [21]. The project ontology consists largely of imported classes from existing mature domain ontologies like ENVO [4,5], the Ontology for Biomedical Investigations (OBI) [2], the Relations Ontology (RO) and other OBO Foundry sources [11].

As a project ontology, ELMO may introduce new terms, but should aim to incorporate those terms into a relevant domain ontology where possible. We used the Simple Standard for Sharing Ontological Mappings [20] to specify entity mappings between modified terms in ELMO and their corresponding terms in existing ontologies to support their semantic (terminological) interoperability, and created comprehensive new term requests for relevant additions (e.g. 204 ecosystem management process terms are under review for inclusion in ENVO).

3 Applications of ELM

3.1 ELM as Building Blocks for a Knowledge Organization System

There are millions of scholarly publications and more being published every year [17]. While paper production increases year over year, the capacity of people to find, read and integrate new science does not keep pace, leading to a growing pool of "dark knowledge" that goes underutilized [12]. Librarians are at the forefront of knowledge organization, but there is great variety in the systems used to organize research and the efficacy of those systems [28]. Rather than proposing a completely novel system, we build upon best practices in digital libraries by using a data schema, linked datasets, ontologies and structured entries to create ELM [31].

The data that is gathered in the format of ELM's schema can be converted into RDF or JSON and used directly by popular search and database technologies. The flexibility of LinkML allows ELM to be used by virtually any repository management software. For instance, a public-facing digital repository could be created using Islandora [14] and made navigable using ELM entries associated with journal DOIs. Alternatively, ELM data could be incorporated into the discovery layer of a library using ExLibris Alma[6], Koha[7] or other integrated library software that supports RDF. It could be incorporated into smaller, project-based repositories built using Islandora[8] or CollectionBuilder[9]. A user could then query for variables they are interested in and associations that occur in study areas similar to their own, providing immediate relevant results (Fig. 3c). Incorporating ELM into ILS discovery layers actively encourages the iterative process of

[5] https://obofoundry.org/.
[6] https://exlibrisgroup.com/.
[7] https://koha-community.org/.
[8] https://islandora.github.io/.
[9] https://collectionbuilder.github.io/.

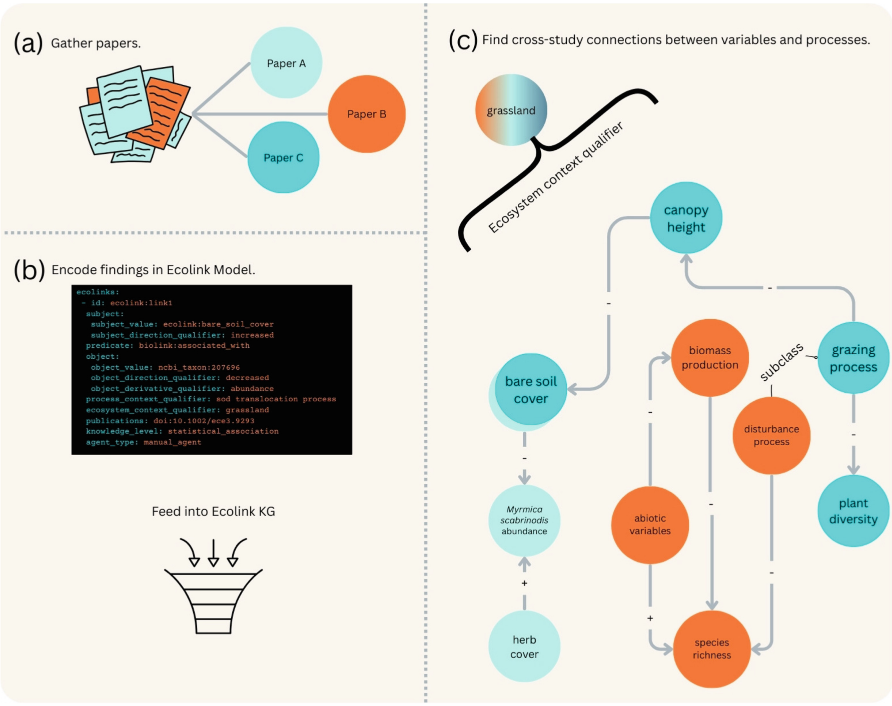

Fig. 3. This diagram shows how the process of encoding papers using the Ecolink Model would function when feeding those papers into a knowledge graph (KG). First, studies are gathered using systematic searching protocols (a). Second, the data is manually or automatically extracted and fitted to the ELM schema in a "semantic parsing" process in which natural language text is translated into an RDF-based syntax using ontology terms, following the structure and constraints in the ELM schema (b). Third, links are established between common node classes (c). Note that (c) is simplified to represent links as + or - for graphical ease. The formal representation would specify whether each node increases or decreases as part of the association.

research, diversifies scholarly conversation, and champions the strategic exploration of sources [7]. In this way, ELM can become the basis for a knowledge organization system that better captures ecological knowledge.

3.2 ELM for Retrieval-Augmented Generation

ELM provides a schema for the creation of structured datasets that describe findings from ecological research, enabling the wider application of artificial intelligence methods. These methods require data that can be collated across ecological gradients, with well-articulated meta-data [18]. Synthesis across ecosystems can improve ecological restoration outcomes and the predictive capacity of restoration ecologists [19]. However, there is a widespread lack of standardization across ecology which hampers interoperability [32]. The combination of ELM and the use of ELMO (and other ontologies) increases the semantic interoperability of knowledge that is represented using the ELM schema, and allows for findings to be collated across ecological gradients.

For instance, retrieval-augmented generation – an artificial intelligence protocol where a large-language model consults a corpus of knowledge to refine its answers – using ecological knowledge may be broadly relevant to the field. This technique may also be useful for building out an ELM knowledge graph based on existing literature [33]. A knowledge graph has the potential to provide information for such an application [26]. ELM would provide a structure for such a graph and could underlie future artificial intelligence research. Additionally, approaches should be explored to add formal semantics to property graphs based on ELM, such as using *OWLStar*[10].

4 Future Research

ELM's model of associations between environmental variables and processes is a powerful, flexible tool for cataloguing ecological knowledge. However, there may be other elements around which core associations could form (e.g. ecosystem types; landscape attributes; climatic regimes). The Biolink Model contains many such associations – gene to gene, phenotype to disease, gene to disease, and so on. Another approach may be making a more precise association type that is focused on containing correlation values, P-values and statistical methods. We provide ELM as a proof of concept to support mobilization of skills and knowledge to use semantic data schema in ecology.

ELM's usefulness will be greatly enhanced by a user interface layer that allows users to query, contribute and edit datasets based on ELM without a significant burden of technical knowledge. The sufficiency and usefulness of ELM should be evaluated by testing with users who need to access the knowledge, such as practitioners of restoration ecology or ecology researchers.

We plan to gather published systematic reviews of ecological literature that document interventions and synthesize them using ELM. This work will also feed the expansion of ELMO – the project ontology – which will in turn improve the comprehensiveness and accuracy of domain ontologies that feed into it.

Finally, we seek to mobilize ELM datasets specifically among conservation and restoration practitioners globally through the establishment of a community of practice. The development of ELM will prioritize accessibility, and we will conduct outreach to communities who may be able to use the data contained within ELM knowledge graphs to justify ecological interventions, manage landscapes and protect biodiversity. Our goal is that ELM, and the knowledge graphs it underpins, will provide an invaluable open knowledge resource globally.

Acknowledgments. We thank Nico Matentzoglu for an abundance of technical support in developing this project, and Nancy Shackelford, Meike Wittmann, Eric Higgs and Stephen Murphy for providing comments.

[10] https://github.com/linkml/owlstar.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alamenciak, T., et al.: Ecological restoration research in Canada: who, what, where, when, why, and how? FACETS **8**, 1–11 (2023). https://doi.org/10.1139/facets-2022-0157
2. Bandrowski, A., et al.: The ontology for biomedical investigations. PLoS ONE **11**(4), e0154556 (2016). https://doi.org/10.1371/journal.pone.0154556
3. Bucur, C.I., Kuhn, T., Ceolin, D., Van Ossenbruggen, J.: Expressing high-level scientific claims with formal semantics. In: Proceedings of the 11th Knowledge Capture Conference, Virtual Event USA, pp. 233–240. ACM (2021). https://doi.org/10.1145/3460210.3493561
4. Buttigieg, P., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E.: The Envo consortium: the environment ontology: contextualising biological and biomedical entities. J. Biomed. Semant. **4**(1), 43 (2013). https://doi.org/10.1186/2041-1480-4-43
5. Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R.L., Mungall, C.J.: The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. J. Biomed. Semant. **7**(1), 57 (2016). https://doi.org/10.1186/s13326-016-0097-6
6. Caufield, J.H., et al.: KG-Hub—building and exchanging biological knowledge graphs. Bioinformatics **39**(7), btad418 (2023). https://doi.org/10.1093/bioinformatics/btad418
7. of College & Research Libraries, A.: Framework for Information Literacy for Higher Education (2016). https://www.ala.org/acrl/standards/ilframework
8. Cunha-Oliveira, T., Ioannidis, J.P.A., Oliveira, P.J.: Best practices for data management and sharing in experimental biomedical research. Physiol. Rev. **104**(3), 1387–1408 (2024). https://doi.org/10.1152/physrev.00043.2023
9. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. Inf. Serv. Use **30**(1–2), 51–56 (2010). https://doi.org/10.3233/ISU-2010-0613
10. Heger, T., et al.: Mapping and assessing the knowledge base of ecological restoration. Restor. Ecol. **32**(8), e13676 (2024). https://doi.org/10.1111/rec.13676
11. Jackson, R., et al.: OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. Database **2021**, baab069 (2021). https://doi.org/10.1093/database/baab069
12. Jeschke, J.M., Lokatis, S., Bartram, I., Tockner, K.: Knowledge in the dark: scientific challenges and ways forward. FACETS **4**(1), 423–441 (2019). https://doi.org/10.1139/facets-2019-0007
13. Jhanwar, M.: Role of assumptions in models: a study. Int. J. Multidisciplinary Innovative Res. (2021)
14. Jones, S., Lampert, C., Lapworth, E., Shaw, S.: Islandora for archival access and discovery. Code4Lib Journal (2023)
15. Jørgensen, S.E., Fath, B.D.: Concepts of modelling. In: Developments in Environmental Modelling, vol. 23, pp. 19–93. Elsevier (2011). https://doi.org/10.1016/B978-0-444-53567-2.00002-8

16. Keith, D.A., et al.: A function-based typology for Earth's ecosystems. Nature **610**(7932), 513–518 (2022). https://doi.org/10.1038/s41586-022-05318-4
17. Khabsa, M., Giles, C.L.: The number of scholarly documents on the public web. PLoS ONE **9**(5), e93949 (2014). https://doi.org/10.1371/journal.pone.0093949
18. Ladouceur, E., Shackelford, N.: The power of data synthesis to shape the future of the restoration community and capacity. Restor. Ecol. **29**(1), e13251 (2021). https://doi.org/10.1111/rec.13251
19. Ladouceur, E., et al.: Knowledge sharing for shared success in the decade on ecosystem restoration. Ecol. Sol. Evid. **3**(1), e12117 (2022). https://doi.org/10.1002/2688-8319.12117
20. Matentzoglu, N., et al.: A Simple Standard for Sharing Ontological Mappings (SSSOM). Database **2022**, baac035 (2022). https://doi.org/10.1093/database/baac035
21. Matentzoglu, N., et al.: Ontology development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies. Database **2022**, baac087 (2022). https://doi.org/10.1093/database/baac087
22. Mazzocchi, F.: Knowledge organization system (kos): an introductory critical account. Knowl. Organ. **45**(1), 54–78 (2018). https://doi.org/10.5771/0943-7444-2018-1-54
23. Moxon, S., et al.: The linked data modeling language (linkml): a general-purpose data modeling framework grounded in machine-readable semantics. CEUR Workshop Proceedings **3073**, 148–151 (2021)
24. Mungall, C.J., et al.: The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. **45**(D1), D712–D722 (2017). https://doi.org/10.1093/nar/gkw1128
25. Palmeri, L., Barausse, A., Jorgensen, S.E.: Ecological Processes Handbook. CRC Press, 0 edn. (2013). https://doi.org/10.1201/b15380
26. Procko, T.T., Ochoa, O.: Graph retrieval-augmented generation for large language models: a survey. In: 2024 Conference on AI, Science, Engineering, and Technology (AIxSET), pp. 166–169. IEEE, Laguna Hills, CA, USA (2024). https://doi.org/10.1109/AIxSET62544.2024.00030
27. Reese, J., et al.: KG-COVID-19: a framework to produce customized knowledge graphs for COVID-19 response (2020). https://doi.org/10.1101/2020.08.17.254839
28. Salatino, A., Aggarwal, T., Mannocci, A., Osborne, F., Motta, E.: A survey on knowledge organization systems of research fields: resources and challenges. Quant. Sci. Stud. 1–37 (2025). https://doi.org/10.1162/qss_a_00363
29. Soergel, D.: Digital libraries and knowledge organization. In: Kruk, S.R., McDaniel, B. (eds.) Semantic Digital Libraries, pp. 9–39. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-85434-0_2
30. Sikos, L.F.: Description Logics in Multimedia Reasoning. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54066-5
31. Soergel, D.: Digital Libraries and Knowledge Organization. In: Kruk, S.R., McDaniel, B. (eds.) Semantic Digital Libraries, pp. 9–39. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-85434-0_2
32. Thessen, A.: Adoption of machine learning techniques in ecology and earth science. One Ecosyst. **1**, e8621 (2016). https://doi.org/10.3897/oneeco.1.e8621
33. Toro, S., et al.: Dynamic retrieval augmented generation of ontologies using artificial intelligence (DRAGON-AI). J. Biomed. Semant. **15**(1), 19 (2024). https://doi.org/10.1186/s13326-024-00320-3, https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-024-00320-3

34. Unni, D.R., et al.: The biomedical data translator consortium: biolink model: a universal schema for knowledge graphs in clinical, biomedical, and translational science. Clin. Transl. Sci. **15**(8), 1848–1855 (2022). https://doi.org/10.1111/cts.13302
35. Wilkinson, M.D., et al.: The FAIR guiding principles for scientific data management and stewardship. Sci. Data **3**(1), 160018 (2016). https://doi.org/10.1038/sdata.2016.18