The Soil Food Web Ontology: aligning trophic groups, processes, and resources to harmonise and automatise soil food web reconstructions

List of authors: Nicolas Le Guillarme[1], Mickael Hedde[2], Anton M. Potapov[3,4,5], Matty P. Berg[6,7], Maria J.I. Briones[8], Irene Calderón-Sanou[1,3], Florine Degrune[2,9], Karin Hohberg[10], Camille Martinez-Almoyna[1], Carlos A. Martínez-Muñoz[11,12], Benjamin Pey[13], David J. Russell[10], Wilfried Thuiller[1]

Affiliations:

1. Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, F-38000 Grenoble, France.
2. Eco&Sols, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
3. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
4. Faculty of Biology, University of Leipzig, Leipzig, Germany
5. Animal Ecology, Johann Friedrich Blumenbach Institut of Zoology and Anthropology, University of Göttingen, Göttingen, Germany.
6. Section Ecology and Evolution, Amsterdam Institute of Life and Environment, Vrije Universiteit, De Boelelaan 1085, 1081, HV, Amsterdam, the Netherlands.
7. Community and Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, University Groningen, Nijenborgh 7, 9747 AG Groningen, the Netherlands
8. Departamento de Ecologia y Biologia Animal, Universidade de Vigo, Vigo, Spain.
9. CIRAD, UMR Eco&Sols, Montpellier, France
10. Soil Zoology Department, Senckenberg - Leibniz Institution for Biodiversity and Earth System Research, Görlitz, Germany.
11. Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt am Main, Germany
12. Zoological Museum, Biodiversity Unit, University of Turku, Turku, Finland
13. LEFE, Univ. Toulouse – CNRS, Toulouse, France.

Corresponding author: Nicolas Le Guillarme; nicolas.leguillarme@univ-grenoble-alpes.fr

Abstract

Although soil ecology has benefited from recent advances in describing the functional and trophic traits of soil organisms, large scale reconstruction of soil food webs is still hampered by the issues that (1) most of the data about trophic interactions and diets are distributed across heterogeneous repositories, (2) no well-established terminology exists for describing trophic groups, feeding processes, and resource types, and (3) large heterogeneity exists in the trophic classification of different soil organisms, or such classifications are lacking. Here, we propose the Soil Food Web Ontology as a new formal framework for representing knowledge on trophic ecology of soil organisms. This ontology captures the semantics of trophic concepts, including consumer-resource interactions, diets and feeding processes, and provides a formalisation of trophic group definitions. This ontology can be used to add semantic annotations to trophic data, thus facilitating the integration of heterogeneous datasets. It also provides lexical resources that can support the development of information extraction tools for the creation of literature-based datasets. Finally, it enables automatic and consistent classification of soil organisms based on their trophic relationships. We argue that, by harmonising the terminology and underlying concepts of soil trophic ecology, our ontology allows a better use of available information on the feeding habits of soil organisms and sounder trophic classifications, thus facilitating the reconstruction of soil food webs and making food web research more accessible, reusable and reproducible.

Keywords: soil food web, trophic groups, ontology, data standardisation, semantic data integration

Introduction

Trophic interactions mediate the majority of the ecosystem services provided by soils, including carbon sequestration, nutrient cycling, pest and pathogen regulation. These soil ecological processes also support plant health and growth, aboveground biodiversity and ultimately ecosystem resilience and stability [1]. Modelling energy and nutrient transfers between soil organisms through accurate reconstructions of soil food webs would provide a better understanding of the relationships between multitrophic assemblages of soil organisms and ecosystem functioning. However, the immense diversity of soil organisms, which spans most of the tree of life, vary in body size over six orders of

magnitude (from 0.1 µm to tens of centimetres) [1], and exhibit a wide range of (micro)habitats and feeding preferences [2], has long hampered the reconstruction of accurate soil food webs.

To deal with this huge diversity in forms and functions, soil biologists resort to classifying belowground biodiversity into guilds of organisms sharing similar niche requirements, or functional groups of organisms acting similarly on a specific ecological function [3]. In particular, when studying multitrophic assemblages, a common approach is to group organisms either into ‘feeding guilds’ of organisms that feed on the same resources, or into ‘trophic groups’ of organisms that both feed on the same resources and have the same consumers [3]. Classifying organisms into relevant trophic groups simplifies food web analysis by allowing the construction of tractable food web models across the whole spectrum of soil organisms [4].

Historically, soil-organism classification into trophic groups was impeded by a persistent lack of consistency in definitions and terminology and by the lack of an overarching framework for classifying all soil biota based on their feeding preferences [3]. In the absence of such a framework, researchers had to resort to user- or clade-specific definitions of trophic groups, which led to heterogeneities in the resolution of food webs and limited our ability to draw generic conclusions across studies [5].

Recent research has worked towards standardisation of trait and trophic group definitions across taxonomic groups. Pey et al. [6] proposed unified definitions for approximately 100 traits and ecological preferences for soil invertebrates, including feeding-related traits, as part of their Thesaurus for Soil Invertebrate Trait-based Approaches (T-SITA). Most recently, Potapov et al. [2] developed an integrative classification of soil consumers from protists to vertebrates. This classification uses a hybrid taxonomic-ecological approach similar to traditional soil food-web models, e.g. Hunt et al. [7], including groups such as Oribatida microbivores, Nematoda fungivores, but uniformly links all group definitions to quantified resource preferences (e.g. plants, algae, litter, bacteria, fungi, dead wood, soil organic matter, other animals) and key traits, including feeding mechanism, body size, vertical stratification and protection traits. Despite not being explicitly included in the classification, potential consumers of each group can be inferred from their traits and thus this classification is conceptually similar to other ‘trophic groups’ classifications as defined by Hedde et al. (see above). This classification provides harmonised definitions of trophic groups across taxonomic groups, and defines a consistent aggregation strategy for food web reconstruction. Yet, to enable automated reconstruction and analysis of soil food webs, we still lack a formal (i.e. standardised and machine-

interpretable) representation of the links between soil organisms' diets (food types), feeding processes, food resources, and trophic groups.

Formal representations of the knowledge in a domain of interest, also called *ontologies*, are increasingly used in biological sciences. Examples include the Extensible Observation Ontology [8], the Biological Collections Ontology, the Environment Ontology, the Population and Community Ontology [9], and the ECOCORE[1] ontology of core ecological entities. Ontologies provide (1) unique identifiers for the concepts (*classes*) and their relationships (*properties*) within a *domain of discourse*, (2) vocabularies of the terms used within this domain, (3) metadata describing the intended meaning of the classes and properties, e.g. textual definitions, and (4) logical *axioms* and definitions that enable specialised softwares (*reasoners*) to automatically access and interpret the knowledge encoded in the ontology [10] (Figure 1). These features make ontologies a powerful tool that facilitates the harmonisation, interpretation, sharing and reuse of domain knowledge, eases the process of integrating heterogeneous data from multiple sources (*semantic data integration*), enables inference of new knowledge using *automated reasoning*, and supports literature-based database creation and curation using *information extraction*.

Concept type	Preferred label (+*synonyms*)	Identifier (IRI)	Textual definition	Logical definition(s)
Diet	zooparasite (*animal parasite)*	SFWO:0000054	A parasite whose host is an animal.	*parasite of* [RO:0002444] <u>some</u> *Metazoa* [NCBITaxon:33208] *capable of* [RO:0002215] <u>some</u> *zooparasitism* [SFWO:0000439]
Trophic process	zooparasitism (*animal parasitism)*	SFWO:0000439	A parasitic process during which a living entity acquires food and energy from an animal host.	*trophic parasitism* [ECOCORE:00000147] <u>and</u> (*has input* [RO:0002233] <u>some</u> *Metazoa* [NCBITaxon:33208])
Trophic group	Nematoda.zooparasites (*zooparasitic nematodes*)	SFWO:0000279		Nematoda.all [SFWO:0000273] <u>and</u> *has member* [RO:0002351] <u>only</u> (*Nematoda* [NCBITaxon:6231] <u>and</u> *zooparasite* [SFWO:0000054])

1 https://bioportal.bioontology.org/ontologies/ECOCORE

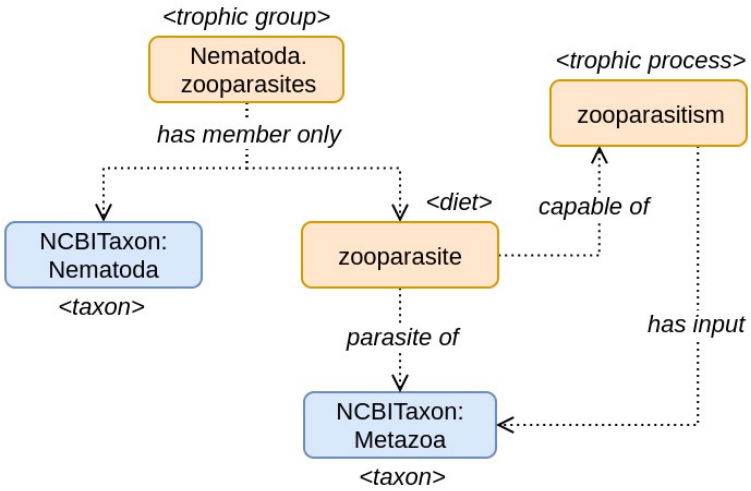

Figure 1. *An ontology is a formal model of a domain of interest as a set of interrelated concepts. SFWO provides unique identifiers, synonyms, human (textual) and machine-interpretable (logical) definitions for a hierarchy of concepts related to the food preferences, food resources, feeding processes, and trophic groups of soil-associated consumers.*

In this paper, we build on previous efforts to harmonise and standardise terms related to the trophic behaviour of soil-associated consumer [2,6] to develop the Soil Food Web Ontology (SFWO), a formal representation of domain knowledge in the field of soil trophic ecology. The development of SFWO pursues several objectives (Figure 2). First, it aims to standardise the terminology in use in soil trophic ecology, e.g. by proposing a reference vocabulary to describe information about diet and feeding processes and harmonising the definitions of trophic groups across taxonomic groups. This will allow to: (1) facilitate knowledge sharing between experts on different taxonomic groups and from different disciplines, (2) facilitate the integration of heterogeneous trophic datasets, (3) support information extraction applications by providing a comprehensive list of terms related to soil trophic ecology, and ultimately to address more complex food-web related research questions in large synthetic studies. Secondly, SFWO aims to provide a logical formalisation of the domain of soil trophic ecology, which will facilitate the creation of food webs through the use of automated reasoning.

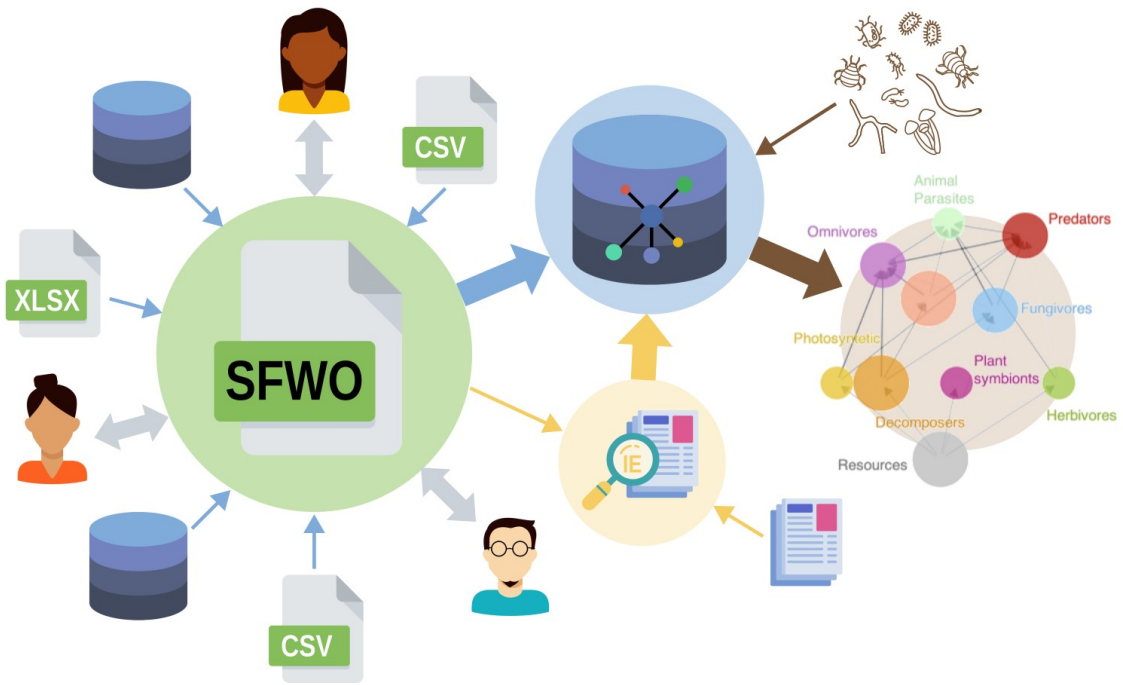

Figure 2. *By providing a standardised terminology and logical formalisation of the domain of soil trophic ecology, SFWO supports knowledge sharing (grey arrows), trophic data integration (blue arrows), extraction of trophic information from literature (yellow arrows), and reconstruction of food webs from soil community data (brown arrows).*

Here, we introduce the Soil Food Web Ontology to encourage the community of (soil) ecologists to adopt SFWO in the assemblage of their trophic data, and to invite experts in the field to contribute to its further improvements. In the following, we provide an overview of the structure and content of the current version of SFWO (v2023-03-09), give some examples of terminological issues that have been addressed by the development team or are still under discussion, briefly describe the development process of SFWO, and finally show the potential of our ontology to support soil food web research by presenting a number of possible applications with some illustrative examples.

Materials and Methods

The Soil Food Web Ontology

SFWO defines a hierarchy of *classes* (concepts) and *properties* (relations). The class hierarchy in an ontology acts much like a taxonomy of concepts, with children (or subclasses) inheriting the properties of their parents (or superclasses). A child class is a specialisation of a parent class: everything about a parent is also true for its children. Object properties are used to assert binary unidirectional relationships between two class instances, or *individuals*. Object properties are organised into their

own hierarchy of properties and subproperties. In SFWO, the minimal definition for a class or property includes a unique *Internationalized Resource Identifier* (IRI) (e.g., http://purl.org/sfwo/SFWO_0000061 is the IRI of the concept *microbivore*), a label (e.g., preferred name for the class/property, in italics throughout the text), and a human-readable textual definition. Most of the classes and properties are associated with a collection of synonyms in different languages, as one of the objectives of SFWO is to provide a comprehensive list of terms related to soil trophic ecology. For ease of handling, it is common practice to refer to IRIs in their abbreviated form, e.g. SFWO:0000061 for *microbivore*. In the following, the IRI of a concept is specified between brackets when this concept is first mentioned or when we feel it necessary. Wherever possible and necessary to allow inference of new knowledge, classes are axiomatized, i.e. logically defined by referring to other classes or properties in the ontology. These axioms, formulated in the *Web Ontology Language* (OWL), provide a machine-interpretable description of the intended meaning of a concept and are particularly useful in keeping SFWO logically consistent and enabling automated reasoning.

The basic core structure of SFWO's class hierarchy consists of three main branches, namely *organismal entities*, *food resources* and *trophic processes* (Figure 3). These three concepts are direct descendants of classes from the Basic Formal Ontology (BFO), an upper-level ontology that provides foundational terms with well-established semantics, facilitating interoperability between ontologies that build on these common foundations. BFO sees reality in terms of a top-level division of all entities into two disjoint categories of *continuant*, e.g. objects, attributes and spatial regions, and *occurrent*, e.g. processes and temporal regions [11].

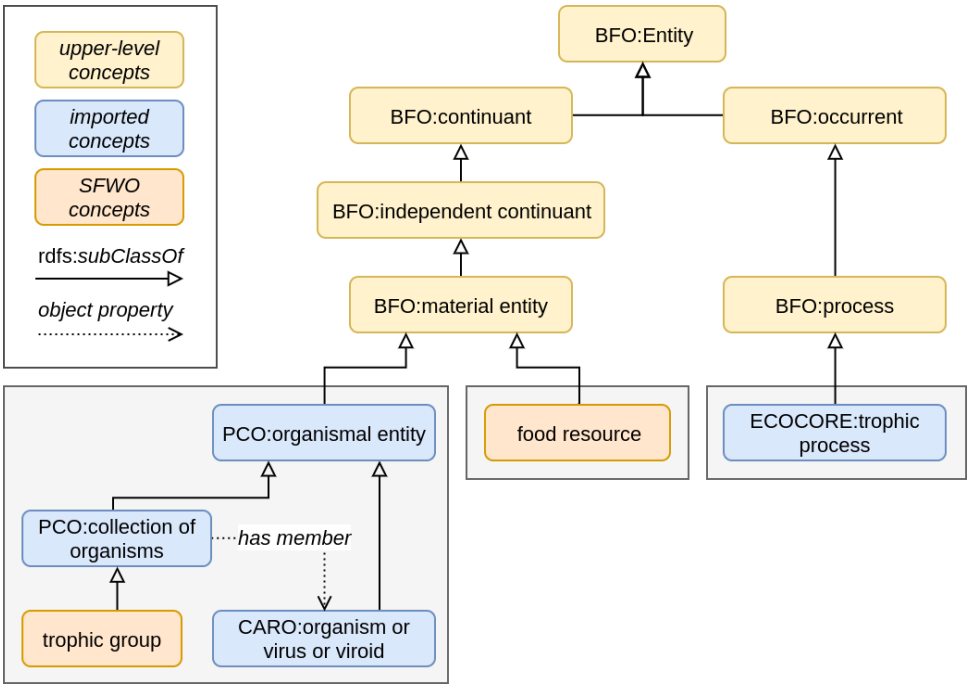

Figure 3. *A high-level view of the three-branch structure of SFWO. SFWO builds upon the abstract concepts defined in BFO (yellow nodes) for interoperability, and reuses as many terms from external ontologies (blue nodes) as possible. It also defines a number of additional concepts (orange nodes) related to diets, food resources, trophic processes and trophic groups. A solid arrow between two classes denotes a subClassOf relation, e.g.* trophic group *is a subclass of* collection of organisms. *Dotted arrows represent semantic relationships between class instances, e.g. a collection of organisms has members instances of the* organism or virus or viroid *class.*

Organismal entities. As the root class of its organismal branch, SFWO reuses the concept of *organismal entity* [PCO:0000031] defined in the Population and Community Ontology (PCO) as "a material entity that is one or more organisms, viruses or viroids". According to PCO, an organismal entity is either an *organism or virus or viroid* [CARO:0001010] (meaning a single organism) or a *collection of organisms* [PCO:0000000]. These two classes define two sub-branches of the organismal branch of the SFWO ontology. The class *organism* or *virus* or *viroid* is the root class for a number of concepts related to taxa, diets and non-taxonomic organisms, whereas the class *collection of organisms* expands into a hierarchy of trophic groups as collections of organisms sharing the same resources and consumers.

Organism. In SFWO, an organism is either (1) a taxon, e.g. Lumbricidae, Insecta, Fungi…, (2) a member of a non-monophyletic grouping of organisms (common names), e.g. algae, protists, microorganisms…, or (3) an organism classified according to its mode of nutrition, e.g. heterotroph, autotroph… (Figure 4).

Taxon. All taxonomic classes are imported from the NCBITaxon ontology[2] which is an automatic translation of the NCBI taxonomy database [12] into OWL and is currently the only taxonomy available in ontological form. To remain as compact as possible, SFWO imports a taxonomic class (and any superclasses in its taxonomic lineage) only if that class is used as part of the logical definition of another class. For instance, the taxonomic class *Nematoda* [NCBITaxon:6231] is used in the logical definitions of the diet *nematophage* [SFWO:0000019] and trophic group *Nematoda.zooparasites* [SFWO:0000279].

Non-monophyletic grouping of organisms. However, some terms describing organismal entities commonly found in soil food web research cannot be associated with a single taxonomic group. This is the case for example of "algae" which is a folk-taxonomic unit representing a large polyphyletic group of diverse photosynthetic organisms that excludes vascular plants and mosses and includes cyanobacteria, seaweed, and single-celled microalgae. Algae are the primary food source of algivorous organisms. SFWO defines a number of such non-taxonomic organismal entities, most of which are needed to axiomatize (i.e. create logical definitions for) diet classes.

Diets. SFWO reuses many nutrition-related classes from ECOCORE. This includes terms that refer to an organism's source of carbon/energy/electron, e.g. *autotroph* [ECOCORE:00000023], *heterotroph* [ECOCORE:00000010] that are all immediate subclasses of the *organism or virus or viroid* class. The hierarchy is then gradually extended with terms referring to different combinations of carbon/energy/electron sources, e.g. *chemoheterotroph* [ECOCORE:00000132], *chemoorganoheterotroph* [ECOCORE:00000127], etc. Among these coarse-grained classes, the *heterotroph* branch is by far the most important and most developed branch of SFWO. This branch contains organismal concepts related to diets (e.g. *nematophage*, *coprophage*, *plant parasite*), feeding mechanisms (e.g., *scraper*, *sucker*, *shredder*), and foraging strategies (e.g., *browser*, *grazer*). Diet terms refer to an organism preference for one or several food resources, e.g., a *saproxylophage* [SFWO:0000012] preference for *dead wood* [SFWO:0000149]*, or a springtail feeder* [SFWO:0000058] preference for *Collembola* [NCBITaxon:30001]. Whenever possible, we give these terms a logical definition in the form of an equivalence axiom. For instance, the class *nematophage*, whose textual

2 https://obofoundry.org/ontology/ncbitaxon.html

definition is “a predator that primarily eats nematodes”, is logically defined using the following equivalent class axiom:

equivalentTo(*nematophage, eats* some *Nematoda*),

which means that any individual involved in an *eats* [RO:0002470] relationship with at least one individual of class *Nematoda* [NCBITaxon:6231] is a *nematophage*. Here, *eats* is an object property (= relationship) imported from the Relations Ontology[3] (RO) and used to assert a trophic interaction between two individuals. The logical definition of *nematophage* broadly matches the textual definition[4]. Logical definitions of diet classes are key to our ability to use ontological reasoning to automatically infer an organism diet(s) and trophic group(s) from its trophic interactions (see subsection on reasoning in SFWO).

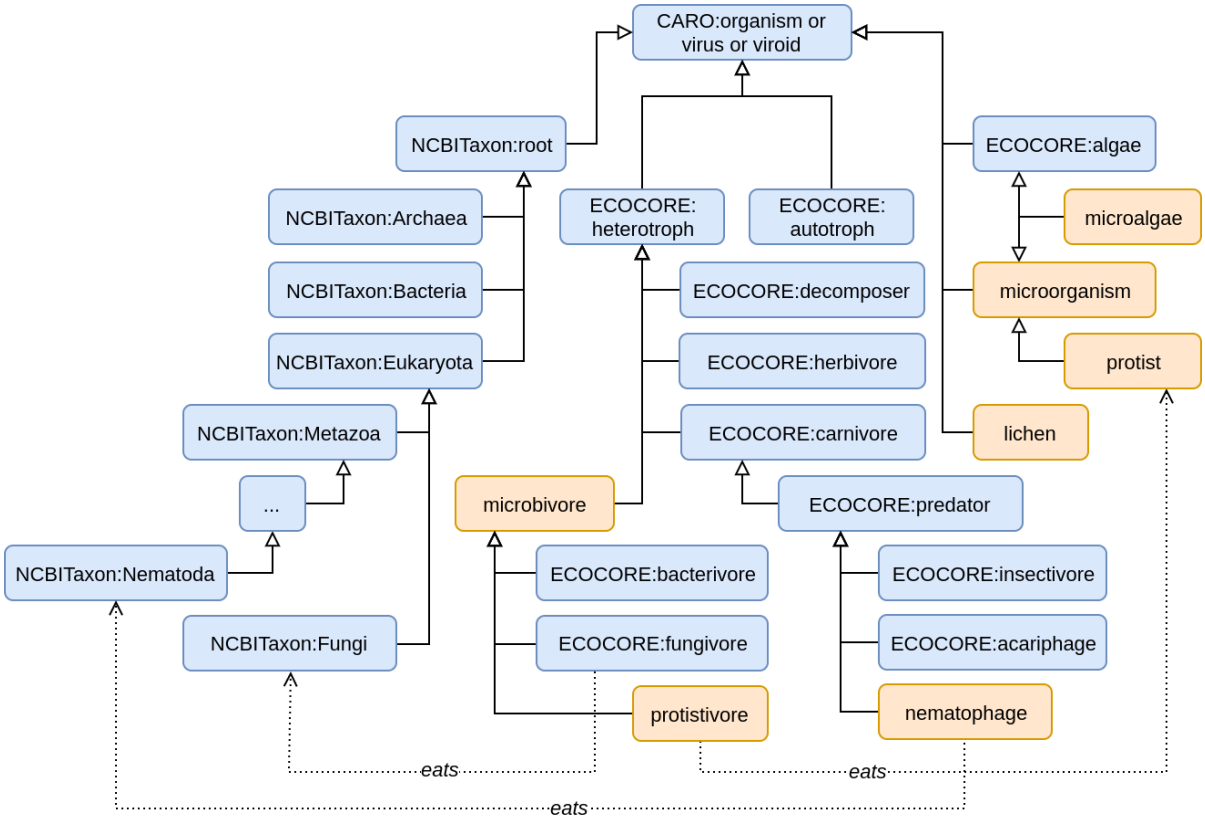

Figure 4. *An excerpt from the organism branch of SFWO. Logical definitions of diet classes create mappings between diets and consumer-resource interactions.*

3 https://github.com/oborel/obo-relations

4 Actually, similarly to textual definitions, logical definitions should preferably adopt the genus-differentia form, so that an equivalence axiom consists of a genus, i.e. the broader class to which the term belongs, and one or more differentia, i.e. the property/ies that distinguish the term from other members of the same class. According to this best practice, the logical definition of *nematophage* should be equivalentTo(*nematophage, predator* and *eats* some *Nematoda*), where *predator* is the genus and *eats* some *Nematoda* is the differentia. In SFWO, we deviate from this best practice for diet classes as it makes inferring an organism diet from the resources it consumes more difficult.

Trophic groups. SFWO introduces a class *trophic group* [SFWO:0000127] defined as “a collection of organisms that feed on the same food resources and have the same consumers” according to the definition given by Hedde et al. in [3]. In the current version of SFWO, this branch of the ontology expands into a hierarchy of trophic groups for soil-associated animals, protists, and fungi.

Trophic groups for protists and animals are taken from the taxonomic-ecological classification of soil-associated consumers proposed by Potapov et al. [2] (Table S1). This hybrid classification distinguishes trophic groups individually within each taxonomic group, which makes it possible to incorporate taxon-dependent trait information in the trophic group definitions. A trophic group is therefore a combination of a taxon, e.g. *Nematoda*, and a (possibly empty) list of traits, including feeding-related traits such as *fungivore*, *herbivore*, *bacterivore*, etc. For illustration, the hierarchy of nematode trophic groups from Potapov et al. is represented in Figure 5.

Fungal trophic groups are taken from the FungalTraits database [13]. Among other functional traits, FungalTraits provides data about the trophic modes (*saprotroph*, *symbiotroph*, *pathotroph*) and guilds (e.g. *litter saprotroph*, *arbuscular mycorrhizal, animal pathogen*) for fungi and fungus-like organisms. Adopting the same approach as for animal and protist trophic groups, we turned FungalTraits trophic modes and feeding guilds into a hierarchy of fungal trophic groups, a subset of which is also shown in Figure 5.

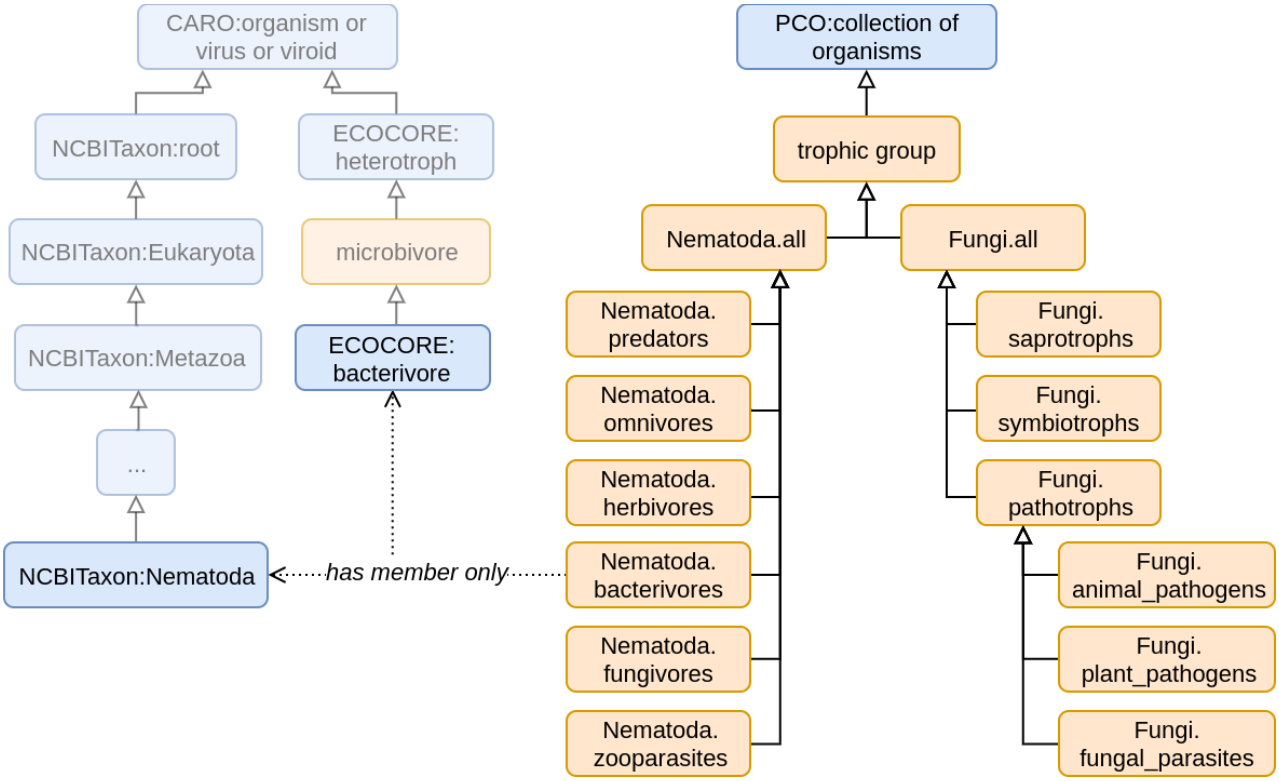

Figure 5. *An excerpt of the hierarchy of trophic groups in SFWO. Trophic group definitions are taken from [2] and [13]. The complete hierarchy consists of 165 trophic group classes for 42 taxa representative of soil biota, from protists to fungi. The axiomatization of trophic group classes links them to taxonomic and diet classes.*

SFWO also provides logical definitions for most of these trophic group classes. These definitions ("axioms") introduce constraints on the type of organisms that can be members of each trophic group. Examples of logical definitions for some nematode trophic groups are given in Table 1. The logical definition of the class *Nematoda.bacterivores* [SFWO:0000275] states that members of this trophic group are necessarily instances of the *Nematoda* and *bacterivore* [ECOCORE:00000098] classes. These logical axioms allow a reasoner to automatically deduce the taxonomic class and diet of organisms from knowledge of the trophic groups of which they are members (see subsection on reasoning in SFWO).

Label	Logical definition
Nematoda.all	*trophic group* [SFWO:0000127] and *has member* [RO:0002351] only *Nematoda* [NCBITaxon:6231]
Nematoda.bacterivores	Nematoda.all [SFWO:0000273] and *has member* [RO:0002351] only (*Nematoda* [NCBITaxon:6231] and *bacterivore* [ECOCORE:00000098])
Nematoda.fungivores	Nematoda.all [SFWO:0000273] and *has member* [RO:0002351] only (*Nematoda* [NCBITaxon:6231] and *fungivore* [ECOCORE:00000015])
Nematoda.zooparasites	Nematoda.all [SFWO:0000273] and *has member* [RO:0002351] only (*Nematoda* [NCBITaxon:6231] and *zooparasite* [SFWO:0000279])

Table 1. *An excerpt of the axiomatization of nematode trophic groups in SFWO, according to the taxonomic-ecological classification of Potapov et al. [2]. The classification is hierarchical, and the root class for a given taxonomic group (here Nematoda.all) is purely taxonomical.*

Food resources. A *food resource* [SFWO:0000128] is defined as "a material entity consumed to provide energetic and nutritional support for an organism". In other words, any entity that is on the right side of a trophic interaction (e.g. *eats*, *parasite of*) is *de facto* a food resource. In SFWO, as in most trophic datasets, this includes (1) taxonomic entities, e.g. *Aphididae* [NCBITaxon:27482] as the primary food source of aphidophagous organisms, (2) non-taxonomic organismal entities, e.g. *algae*, *protist*, (3) organism parts, e.g. *pollen* [PO:0025281], *blood* [UBERON:0000178], *mycelium* [FAO:0000011], and (4) environmental material, e.g. *dead plant material* [SFWO:0000146], *soil organic matter* [ENVO:04000008], etc. Being a food resource is not an inherent characteristic of an entity (e.g., a collembola, a piece of dead wood, a plant or a leaf), but the consequence of this entity being the object of a trophic interaction. As a result, the ontology classes describing food resources, most of which are imported from specialised ontologies, e.g. the Plant Ontology (PO) for plant parts (*leaves*, *seeds*, *roots…*), are scattered over different branches of the ontology. However, we decided to create a convenience class *food resource* to gather the entities that are commonly considered as trophic resources in the literature or food-web reconstruction under a common root concept. This could be useful for text mining applications (e.g. extracting mentions of food resources from the literature as part of a trophic interaction extraction pipeline) and for trophic database exploration. Note that taxonomic classes are not included in this food resource branch, as taxonomic entities all come from the same external ontology (NCBITaxon) and already have their own branch in SFWO.

Trophic processes. Mirroring the hierarchy of diet classes, SFWO provides a hierarchy of terms referring to trophic processes under the root term *trophic process* [ECOCORE:00000008] (see Figure 3). For the sake of completeness, all the nutrition-related classes have their processual counterparts, e.g. coprophagy [SFWO:0000043] (a process) is the trophic process involving a *coprophage* [ECOCORE:00000176] (a diet) feeding on *faeces* [UBERON:0001988] (a resource). Although mostly useful for building a comprehensive list of terms related to trophic ecology, these process classes are also used to axiomatize the corresponding diet classes (see the example with *zooparasite* and *zooparasitism* in Figure 1). These logical axioms use the object property *capable of* [RO:0002215] that defines a relation between an entity (e.g. an organism) and a process the entity has the ability to carry out. This results in a complete 1-to-1 mapping between diets and trophic processes.

Trophic interactions. Trophic interactions are relational concepts linking a consumer to a food source. They are naturally represented in SFWO using object properties. Following best practices of ontology development, SFWO reuses object properties from the Relations Ontology (RO). In particular, SFWO imports an ecological subset of RO whose top-level object property *ecologically related to* [RO:0002321] is the root of a branch that comprises many terms for representing biotic and trophic interactions between individuals. The two terms most often used as part of the diet class logical definitions are *eats* [RO:0002470] — e.g. in the definition of *nematophage* — and *parasite of* [RO:0002444] — e.g. in the definition of *zooparasite*. Note that *parasite of* is not considered to be a trophic interaction but a symbiotic interaction in RO, this latter being defined as “a biotic interaction in which the two organisms live together in more or less intimate association”. In its current version, RO is missing terms to represent symbiotrophic interactions in which one organism acquires nutrients through a symbiotic relationship with another organism, e.g. trophic parasitism, mycorrhizal associations, etc. The object property *symbiotrophically interacts with* [SFWO:0000099] was created in SFWO as a temporary solution, and a new term request has been submitted to the GitHub issue tracker of RO[5].

5 https://github.com/oborel/obo-relations/issues/647

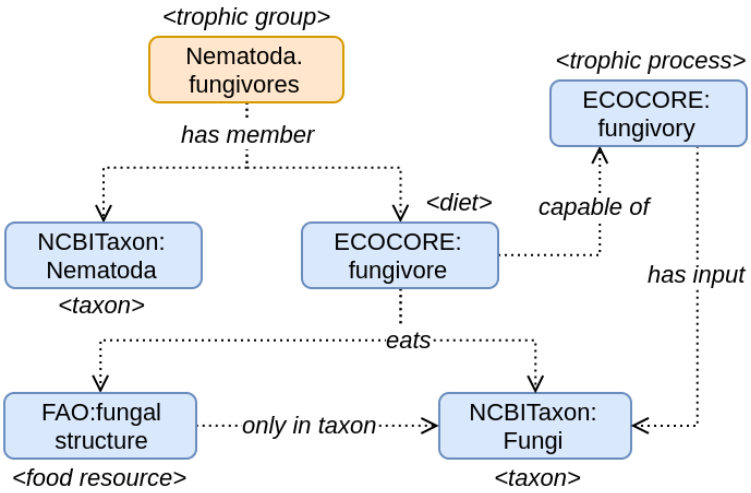

Figure 6. *Class axiomatization in SFWO links diet, resource, trophic process, and trophic group concepts together, and makes it possible to derive additional knowledge using inference.*

Reasoning in SFWO

Axiomatized class definitions in SFWO create semantic links between taxonomic entities, trophic groups, diets, food resources and trophic processes (Figure 6), and allow a reasoner to automatically classify individuals using ontology-based reasoning. In particular, this allows for three types of inference:

1. inference of a consumer's diet(s) from consumer-resource interactions;
2. inference of a consumer's (potential) trophic interactions from its diet(s);
3. inference of a consumer's taxonomic class and (potential) diet(s) from its hybrid taxo-trophic group(s).

In addition to logical definitions of trophic concepts, SFWO includes a number of general class axioms, i.e. logical axioms with anonymous classes on the left-hand side, to link organisms to trophic groups. For instance,

subClassOf(*Nematoda* and *fungivore, member of* value *Nematoda.fungivores*)

is a general class axiom stating that any individual organism that is an instance of the anonymous class (*Nematoda* and *fungivore*), i.e. the intersection of classes *Nematoda* and *fungivore* [ECOCORE:00000015], is also a member of the trophic group *Nematoda.fungivores* [SFWO:0000276]. Such general class axioms allow for a fourth type of inference

4. inference of a consumer's trophic group(s) from its taxonomic class and diet(s).

An example of inference is given in Figure 7. In this example, the fact that the nematode *Aphelenchus avenae* feeds on hyphae is asserted. The predicate *rdf:type* is used to assert that an individual is an instance of a specific class. Here, i_1 is an instance of *A. avenae* [NCBITaxon:70226] and i_2 is an instance of hypha [FAO:0001001]. Given information of a trophic interaction between i_1 and i_2, a reasoner is able to derive, using SFWO's axioms and inference rules, the following facts:

1. i_1 is a *Nematoda*, because *A. avenae* is a subclass of *Aphelenchus*, which is a subclass of *Aphelenchidae*, which is a subclass of... , which is a subclass of *Nematoda* ;
2. i_2 is a *fungal structure*, because *hypha* is a subclass of *multicellular fungal structure*, which is a subclass of *fungal structure* ;
3. i_1 is a *fungivore*, because i_1 eats i_2 and i_2 is a *fungal structure* and *fungivore* is logically equivalent to *eats* <u>some</u> *(Fungi* <u>or</u> *fungal structure*) ;
4. i_1 is a member of the trophic group *Nematoda.fungivores,* from inferred facts 1 and 3 and the general class axiom linking *Nematoda*, *fungivore*, and *Nematoda.fungivores*.

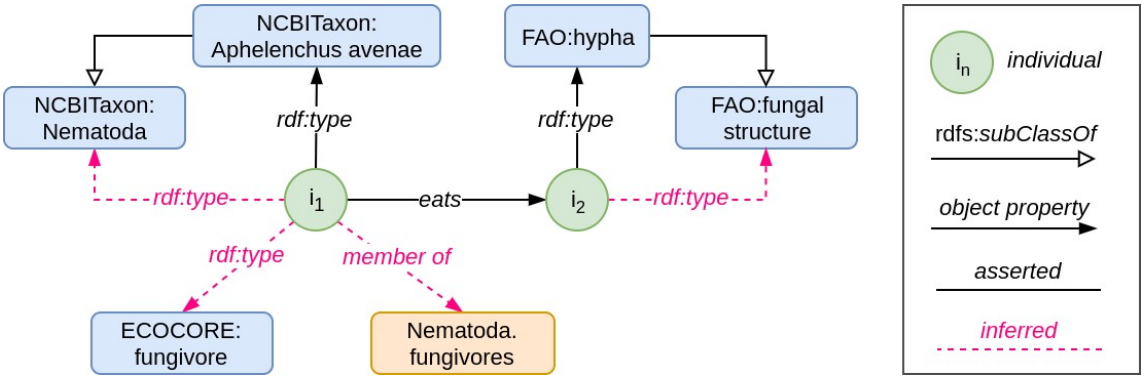

Figure 7. *A reasoner is able to derive logical consequences from asserted facts using the ontology axioms. Here, the diet and trophic group of A. avenae are inferred from the asserted fact that this species of nematode feeds on hyphae.*

Terminological issues

The development of SFWO has highlighted a number of terminological issues currently present in soil ecology that were, at least partly, addressed here (see terminological discussions on SFWO's GitHub issue tracker[6]). Below we describe a few of these issues as examples of the challenges in harmonising the terminology in a scientific domain and modelling these concepts as part of a formal

6 https://github.com/nleguillarme/soil_food_web_ontology/issues

representation of the domain. In our decisions we tried to adhere to the most widely accepted use of terms, rather than creating a *de-novo* (perfectly consistent) system.

Among the most evident issues was the terminological diversity of diets/trophic processes in soil food webs. For example, organisms feeding on fungi are named in different literature sources *fungivores* (e.g. [14]), *mycophages*, or *mycotrophs* (e.g. [15]). Following the literature and expert exchange, we uniformly used *-vore* and *-phage* as synonyms for organisms that feed on specific food resource by consuming the whole or part of it (*herbivore* = *phytophage*, *bacterivore* = *bacteriophage*); *-trophy* is commonly used as a broader term depicting any type of energy/nutrient acquisition from specific food resource, including for instance symbiotic exchanges (*symbiotrophs* [ECOCORE:00000045], such as *mycorrhizal* [SFWO:0000072] fungi) or extracellular degradation and intake of organic matter (microbial decomposition process, implemented by *saprotrophs* [ECOCORE:00000056]). At present, the Relations Ontology does not provide the necessary properties to distinguish between internal and external digestion. As a workaround, SFWO sometimes introduces a taxonomic constraint in logical definitions to differentiate between -vore/-phage (used for animals) and -troph concepts (used for microbes).

Another (open) terminological issue deals with the term *algae* which refers to a polyphyletic group of both macro and microscopic photosynthetic organisms. In soil science, however, the term *algae* is used as a convenient synonym of microalgae, as all algae in soil are unicellular. In this context, it is problematic to reuse the term *algivore* from ECOCORE, as this ontology considers algivores to be a subclass of herbivores, whereas soil ecologists generally consider (micro)algivores to be microbivores, i.e. feeding on microorganisms, such as bacteria, fungi, and/or protists – the latter comprising many soil microalgae [16]. Instead, we decided to create the concepts *microalgae* [SFWO:0000464] as a subclass of *microorganism* [SFWO:0000062] (adding algae as a broad synonym of *microalgae*) and *microalgivore* [SFWO:0000465] as a subclass of *microbivore* [SFWO:0000061]. This better reflects the common understanding of these terms in soil ecology, as well as the fact that soil algae are selectively consumed mainly by bacterial, fungal, and protist feeders, rather than vascular plant herbivores, e.g. root-feeding nematodes, curculionid larvae [2]. However, this decision is suboptimal as we probably need an easy term to depict all trophic processes related to primary production [17].

We also discussed the scope and classification of *detritus*. Moore et al. [18] defined detritus based on Swift et al. [19] as "... any form of non-living organic matter, including different types of plant tissue (e.g. leaf litter, dead wood, aquatic macrophytes, algae), animal tissue (carrion), dead microbes, faeces (manure, dung, faecal pellets, guano, frass), as well as products secreted, excreted or exuded from organisms (e.g. extracellular polymers, nectar, root exudates and leachates, dissolved organic matter, extracellular matrix, mucilage)." This broad definition includes nectar and root exudates, and as such we should classify rhizosphere microorganisms, nectar feeders (e.g. bees), and any body surface-dwelling symbionts as *detritivores* or *saprotrophs*. Nectar and exudates are connected to the surface of a living organisms and, at present in SFWO, we classify *nectarivore* as a subclass of *herbivore* (i.e. feeding on living plant or macroalgal matter), because the opposite would consequently require classifying numerous nectar-feeding insects as detritivores if the ontology was to be be expanded to aboveground food webs.

Another topic of discussion was the nature of the *predation* concept. It was argued that the concept of predation does not refer to a specific diet characterised by the consumption of a particular type of food, but rather to a food acquisition strategy. Similarly to ECOCORE, from which we reuse many classes including the *predation* [ECOCORE:00000102] class, SFWO does not make a clear distinction between feeding preferences (e.g. *insectivore*, *nematophage*, *herbivore*) and feeding strategies (e.g. *predator*, *filter-feeder*, *grazer*). In addition, the term predator is very commonly used in the soil literature to refer to animals or protists feeding on animal/protist prey. We decided to remain with this widely accepted usage. However, this may be worthy of further discussion, which could include other such terms, e.g. parasite.

A final example of a non-trivial terminological (and ontological) issue concerns the place of *saproxylophagy* in the hierarchy of concepts. Saproxylophagy is the process of eating dead or decaying wood, and as such, it is a specialisation (subclass) of xylophagy, which describes the feeding activities of an herbivorous animal whose diet consists primarily of wood (living or dead). However, this would lead to the conclusion that saproxylophagy is a form of herbivory, when in fact it is most commonly considered as a form of detritivory. We have chosen to make *saproxylophage* [SFWO:0000025] a subclass of *phytosaprophage* [SFWO:0000025], itself a subclass of *detritivore*, to reflect this common use of the term in the literature.

SFWO's development and maintenance

The ontology design process started with a small team of experts in soil invertebrate ecology who provided a list of over 100 terms related to primary food groups, diets and food resources of soil-associated consumers, their definitions and relationships between diets and food resources, e.g. *entomophage* → *Insecta* or *saproxylophage* → *dead wood*. A subset of these core terms was manually aligned with classes from existing ontologies using Ontobee's search engine [20]. For instance, the term "bacterivore" matched the class *bacterivore* [ECOCORE:00000098] defined in the ECOCORE ontology. These classes were used as "seeds" for module extraction, a popular strategy for ontology reuse that avoids the overheads involved in importing complete ontologies [21]. Modules were merged to form the backbone of SFWO. New classes for the terms that could not be imported from existing ontologies were created manually using Protégé[7], a free and open-source ontology editor. The ontology development workflow is managed and automated using the Ontology Development Kit [22]. The ontology is under version control in a GitHub repository[8] which is also used to track changes and issues, and manage releases.

We tried to make SFWO conform to the shared principles for developing ontologies, as prescribed by the Open Biomedical Ontologies (OBO) Foundry [23], including the use of BFO as an upper-level ontology, and the use of object properties from the Relations Ontology. This facilitates both interoperability between SFWO and OBO ontologies and terms reuse. In particular, we reuse many nutrition-related terms from the ECOCORE ontology, organismal terms from the Population and Community Ontology, anatomical/metabolic terms from the Plant Ontology, the Fungal Anatomy Ontology and UBERON to describe plant/fungal/animal resources, and taxonomic classes from the NCBITaxon ontology.

SFWO is continuously updated, improved and enriched thanks to contributors' suggestions for revisions. Currently, SFWO brings together a community of a dozen soil ecology experts. Contributors are invited to suggest and discuss changes in the ontology using the issue tracker of the project's GitHub repository. Requests for changes/modifications are regularly submitted to the ontology steering committee[9] for discussion and validation. If approved, changes are applied using the above methodology (extraction of modules or creation of new classes, manual editing to add synonyms or to

7 https://protege.stanford.edu/
8 http://github.com/nleguillarme/soil_food_web_ontology
9 The composition of the steering committee can be found on the project's GitHub page.

modify textual or logical definitions...) and the revised version is released. A release of SFWO consists of a single OWL file (sfwo.owl). The latest version of the ontology is always accessible at the following URL: http://github.com/nleguillarme/soil_food_web_ontology/releases/latest/download/sfwo.owl.

Results

SFWO is a community-led ontology with a double purpose: to be a reference ontology (including a comprehensive vocabulary) for soil trophic ecology, and to be a resource for ontology-driven applications in this domain. SFWO was specifically designed to meet four main users' needs: trophic trait dataset standardisation, semantic data integration, automated food web reconstruction, and trophic information extraction from text.

Standardisation of trophic trait datasets

The lack of standardisation for trait data, including feeding-related trait data, hampers the efficient reuse and synthesis of published datasets [24]. A major challenge in trait-data standardisation is the lack of widely accepted and unambiguous trait definitions. The T-SITA thesaurus [6] is an effort to provide such definitions for soil invertebrates. Out of many non-trophic traits, T-SITA defines a hierarchy of 20+ diet-related terms. SFWO goes one step further, with an ever-increasing vocabulary of over 500 terms (including synonyms) for describing trophic interactions, diets, food resources and trophic processes. SFWO has the potential to become the consensus terminology for describing, storing and publishing standardised soil trophic data. This assures data providers that their data is unambiguous, interoperable with other datasets using the same standard vocabulary, and more easily comparable/reusable as part of future data aggregation initiatives and synthetic studies.

We demonstrate the applicability of SFWO to standardisation of trophic datasets by adding semantic annotations to two open datasets, namely the BETSI database of soil invertebrate traits [25] and the Hexapoda dietary trait dataset from Rainford and Mayhew [26]. Table 2 and Table 3 show a few samples of these datasets, enriched with the IRIs of the concepts in SFWO that best capture the semantics of the data. This makes the two datasets easier to interpret and integrate, as they now use the same unambiguous vocabulary to describe dietary information.

taxon_name	trait_name	attribute_trait	attribute_trait_iri
Abacetus (Astigis)	Diet	Zoophage	ECOCORE:00000088 (*carnivore*)
Abax (Abax) ovalis (Duftschmid, 1812)	Diet	Zoophage	ECOCORE:00000088 (*carnivore*)
Abax (Abax) ovalis (Duftschmid, 1812)	Diet	Necrophagous	ECOCORE:00000090 (*scavenger*)

Table 2. *Samples of the BETSI database of soil invertebrate traits [25]. These samples consist of dietary information for one genus and one species of Carabidae. Data in the attribute_trait column are matched with the corresponding concepts in SFWO.*

Taxon	Larval diet	Adult diet	larval_diet_iri	adult_diet_iri
Coleoptera Trachypachidae	4	4	ECOCORE:00000089 (*predator*)	ECOCORE:00000089 (*predator*)
Coleoptera Cupedidae	1	2	ECOCORE:00000015 (*fungivore*)	ECOCORE:00000055 (*detritivore*)
Coleoptera Micromalthidae	1	1	ECOCORE:00000015 (*fungivore*)	ECOCORE:00000015 (*fungivore*)

Table 3. *Samples of the Hexapoda dietary trait dataset from [26]. These samples show information about the diet of three families of beetles at the larval and adult stage. Dietary information is encoded using a numerical encoding scheme: 1=Fungivory, 2=Detritivory, 4=Predation. Data in the* Larval diet *and* Adult diet *columns are matched with the corresponding concepts in SFWO.*

Trophic data integration and knowledge base construction

Ecology has entered the Big Data realm for a few years now, yet large-scale studies of ecological systems is still hampered by the fact that biodiversity data are highly heterogeneous and widely distributed, although increasingly available. Semantic data integration, i.e. the process of combining data from multiple sources into a single knowledge base using ontologies, has the potential to facilitate large synthetic ecological studies by (1) providing the user with the ability to seamlessly manipulate data from disparate sources regardless of the original format or location of the data, (2) making heterogeneous data interoperable as part of a single knowledge base using ontologies as a mediating representation, and (3) providing data with semantic context ("meaning") to support its interpretation. SFWO is a major contribution for trophic data integration, as it provides the semantic foundations for representing knowledge in the domain of soil trophic ecology, across taxonomic groups and trophic levels. This makes SFWO a prime choice as a conceptual model to reconcile schematic and semantic heterogeneities between data sources and build a unified knowledge base of

trophic data for all soil biota [27] (Listing 1). Such a knowledge base would be a valuable resource for multitrophic studies of soil ecosystems and soil food web reconstruction, as it would provide a unified access to multigroup, multitrophic information.

subject	*predicate*	*object*	
_:betsi_carabidae_0	rdf:type	CARO:0001010	(1)
_:betsi_carabidae_0	rdf:type	NCBITaxon:170239	(2)
_:betsi_carabidae_0	rdf:type	ECOCORE:00000088	(3)
_:betsi_carabidae_1	rdf:type	CARO:0001010	
_:betsi_carabidae_1	rdf:type	NCBITaxon:106379	
_:betsi_carabidae_1	rdf:type	ECOCORE:00000088	
_:betsi_carabidae_2	rdf:type	CARO:0001010	
_:betsi_carabidae_2	rdf:type	NCBITaxon:106379	
_:betsi_carabidae_2	rdf:type	ECOCORE:00000090	
_:rainford_mayhew_0	rdf:type	CARO:0001010	
_:rainford_mayhew_0	rdf:type	NCBITaxon:60838	
_:rainford_mayhew_0	rdf:type	ECOCORE:00000089	
_:rainford_mayhew_1	rdf:type	CARO:0001010	
_:rainford_mayhew_1	rdf:type	NCBITaxon:50548	
_:rainford_mayhew_1	rdf:type	ECOCORE:00000055	
_:rainford_mayhew_2	rdf:type	CARO:0001010	
_:rainford_mayhew_2	rdf:type	NCBITaxon:110040	
_:rainford_mayhew_2	rdf:type	ECOCORE:00000015	

The identifier of the individual described in the first line of Table 2

(1) This triple expresses the fact that _:betsi_carabidae_0 is an instance of *organism or virus or viroid* [CARO:0001010]

(2) This triple expresses the fact that _:betsi_carabidae_0 is an instance of *Abacetus* [NCBITaxon:170239]

(3) This triple expresses the fact that _:betsi_carabidae_0 is an instance of *carnivore* [ECOCORE:00000088]

Listing 1. *A semantic data integration approach such as the one described in [27] can transform data from multiple heterogeneous data sources into a common format — here, the Resource Description Framework (RDF) data model. The atomic unit of data in RDF is called a triple. A RDF triple expresses a relationship between two entities in the form of a subject-predicate-object expression. In this example, the annotated data in Table 2 and Table 3 have been transformed to RDF triples and combined into a single RDF dataset. This provides a unified view over the initial data sources that can be queried using the SPARQL Query Language[10] for RDF to retrieve information across sources.*

Soil food web reconstruction

Several aspects of SFWO can support multitrophic studies and soil food web reconstruction from community composition data. First, SFWO contributes to fill the “Eltonian shortfall” [28] in soil ecology, i.e. a lack of knowledge about the inter- and intra-specific relationships within soil communities, by facilitating the integration of trophic data for different taxonomic groups from multiple data sources, making trophic knowledge easier to access, and making implicit knowledge explicit using automated reasoning. Secondly, with over 160 trophic group classes (most of which axiomatized) covering a

10 https://www.w3.org/TR/rdf-sparql-query/

broad spectrum of taxonomic groups, SFWO provides a common, logically-sound framework for soil-associated consumers classification. This framework solves many issues related to inconsistent definitions of trophic groups across published studies and the heterogeneity in food web aggregation levels that they entail, thus making food webs reconstructed in different research contexts more reusable, comparable and combinable as part of e.g., meta-analyses. In addition, SFWO's logical axioms make it possible to automate the process of assigning taxa to trophic groups using a reasoner, thus reducing the burden of manual trophic group assignment.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX RO: <http://purl.obolibrary.org/obo/RO_>
PREFIX NCBITaxon: <http://purl.obolibrary.org/obo/NCBITaxon_>
PREFIX SFWO: <http://purl.org/sfwo/SFWO_>
SELECT ?taxon ?taxon_name
WHERE {
  ?consumer  RO:0002350     SFWO:0000369.   (1)
  ?consumer  rdf:type       ?taxon.         } (2)
  ?taxon     rdfs:subClassOf NCBITaxon:1.   }
  ?taxon     rdfs:label     ?taxon_name.    (3)
}
```

(1) Assign all individuals that are *member of* [RO:0002350] the trophic group *Coleoptera.detritivores* [SFWO:0000369] to variable ?consumer

(2) Assign the taxonomic classes of all the individuals obtained with (1) to variable ?taxon; NCBITaxon:1 is the root class of the NCBITaxon ontology

(3) Assign the labels of the taxonomic classes obtained with (2) to variable ?taxon_name

taxon	taxon_name
NCBITaxon:106379	Abax ovalis
NCBITaxon:50548	Cupedidae

Figure 8. *Information in RDF databases can be retrieved using SPARQL queries. A SPARQL query starts with the declaration of a number of prefixes which enable writing prefixed names in queries instead of having to use full IRIs. The SELECT clause identifies the variables to appear in the query results, and the WHERE clause provides the pattern to match against the data. A SPARQL variable starts with an ?. This query retrieves the IRIs (stored in the variable ?taxon) and names (stored in the variable ?taxon_name) of all the taxa that are members of the trophic group Coleoptera.detritivores. When run against the data in Listing 1, this query returns two taxa, namely Abax ovalis* [NCBITaxon:106379] *and Cupedidae* [NCBITaxon:50548].

The query in Figure 8 exemplifies the benefits of inference for assigning taxa to trophic groups. The query retrieves all the taxa that are members of the trophic group *Coleoptera.detritivores*. Evaluated against the data in Listing 1, the query returns two results: (NCBITaxon:106379, *Abax ovalis*) and (NCBITaxon:50548, *Cupedidae*). This is the consequence of a reasoner inferring that *A. ovalis* and *Cupedidae* are both indirect subclasses of *Coleoptera*, and deducing from the necrophagous habits of *A. ovalis*, available from the BETSI database (Table 2), that this species is a detritivore since *necrophage* [ECOCORE:00000090] is a subclass of *detritivore* in SFWO. The fact that *Cupedidae* are detritivores is stated explicitly in Rainford and Mayhew's database (Table 3). Finally, the reasoner

combines these intermediate results to infer that *A. ovalis* and *Cupedidae* are both members of the trophic group *Coleoptera.detritivores*.

Trophic information extraction

The above applications of SFWO rely on our ability to compile large databases on the trophic ecology of soil consumers. Although much data is available from previously published research, the creation of literature-based datasets requires significant manual investment for literature searching, acquisition, screening, data extraction, and harmonisation of entities, such as species or trait names [29]. An alternative is to use information extraction approaches to automatically turn unstructured text into structured data, an approach currently taken by the Specialised Information Service Biodiversity Research[11] (BIOfid), which aims at extracting structured data from legacy Central European literature through semantic role labelling. With its large and ever-increasing terminology related to soil trophic ecology, SFWO is a valuable lexical resource for trophic information extraction. The most straightforward way to use SFWO as part of an information extraction pipeline is to assemble all or some of the terms defined in the ontology into a gazetteer, i.e. a list containing names of entities of interest, and use this gazetteer to find mentions of these entities in text (Figure 8). SFWO is particularly suitable to detect mentions of trophic processes, food resources and diets in text. In combination with a taxonomic entity recognition system such as the one described in Le Guillarme and Thuiller [30], SFWO could support the automatic creation, curation and updating of literature-based trophic trait datasets.

11 www.biofid.de

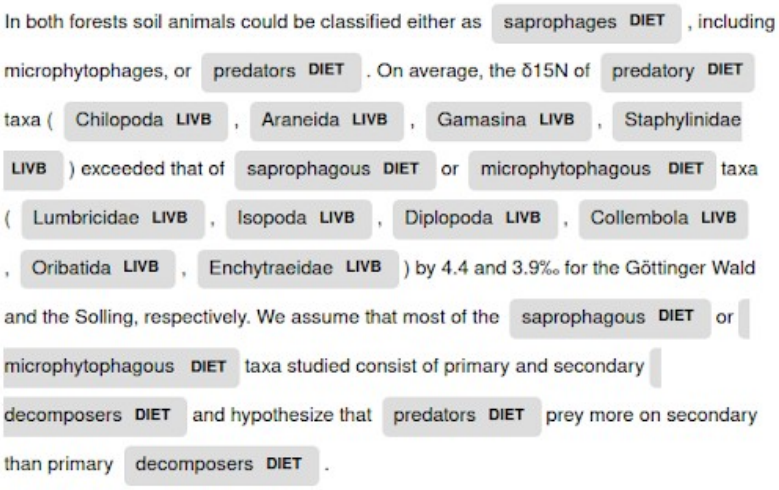

Figure 8. *SFWO can be used as a dictionary of terms (or gazetteer) for the automatic retrieval of mentions of feeding behaviours in the literature. In this example, all of the terms in SFWO terms are grouped into a single entity class (DIET). The taxonomic entities (LIVB) are extracted using TaxoNERD [30].*

Discussion and conclusions

The Soil Food Web Ontology is a collaborative and ongoing effort at developing a standard conceptual model of soil trophic ecology that is both interpretable by humans and computers. Building on existing reference ontologies, continuously curated and enriched by a dozen soil scientists in constant dialogue with an ontology engineer, SFWO is an unprecedented resource for better knowledge management in the field. Opening many avenues for ontology-based applications, including trophic trait datasets standardisation and integration, computer-assisted food web reconstruction, literature-based database creation/curation…, SFWO has the potential to assist soil scientists in collecting, annotating, interpreting, (re)using and sharing trophic data.

In its current state, SFWO is the most comprehensive conceptual model (and vocabulary) for describing trophic knowledge related to soil biota. However, SFWO is likely missing some trophic concepts and many synonyms. In the next development iterations, we plan to add more synonyms, with an increased effort in adding them in multiple languages. We will pursue the axiomatization of trophic group classes, the definition of some of them involving non-trophic traits. For instance, the group *Lumbricina.all* is subdivided in three subgroups: *Lumbricina.epigeic, Lumbricina.anecic, Lumbricina.endogeic* based on vertical stratification of microhabitat. Providing logical definitions for these classes requires incorporating the concepts of epigeic, anecic, and endogeic organisms which are not trophic concepts *per se*, although closely related to feeding behaviours. We also plan to refine

and enrich the classification in trophic groups, particularly with regard to bacteria which are currently absent. In the longer term, we aim to increase the expressiveness of SFWO so that it makes it easier to take into account the plasticity of feeding behaviour in certain taxonomic groups. A first step in this direction was the introduction of life stage concepts in the ontology. We would like to explore the possibility of introducing representational mechanisms that would help quantify uncertainty in diet assignment, resource preferences, and the strength of trophic interactions.

Perhaps the most difficult aspect of developing an ontology is getting it accepted and adopted by the scientific community. A few factors can facilitate engagement of the community of soil scientists. First, the ontology must be consensual, which means that it must capture the knowledge of the domain in a way that is accepted by the majority, if not by all. To help build and maintain consensus over the years, we have designed SFWO's development as a collaborative process in which everybody can participate, along the lines of open source software development. Therefore, on behalf of the steering committee, we invite soil scientists, data providers, and anyone interested in using SFWO to contribute to the improvement of the ontology through the project's GitHub repository. Furthermore, to facilitate the adoption of SFWO and the dialogue between ontology developers and domain experts (likely unfamiliar with the arcana of ontological languages such as OWL), we have made the ontology available in more user-friendly formats. A browsable version of SFWO is available online on AgroPortal[12], with the IRIs of the SFWO concepts all being redirected to AgroPortal. It also includes a graphical representation of the concepts hierarchy, accessible from the "Visualization" tab in the Classes view. In addition, each release of SFWO will be accompanied by an export of the ontology in a tabular format (CSV file). The export file of the current version of SFWO accompanying this paper is provided as Supplementary Material. Finally, we will provide comprehensive documentation so that users can better understand the architecture of the ontology and the design choices that led to that architecture.

Eisenhauer et al. [31] identified better data management, including data standardisation and integration, as one of the prominent research frontiers in soil ecology. This requires data standards (from the less expressive controlled vocabularies to the more expressive ontologies) to aid data harmonisation and compilation. Our Soil Food Web Ontology is made to tackle this issue and provide the semantic resources needed to enable researchers to make best use of the existing and future

12 http://agroportal.lirmm.fr/ontologies/SFWO

data on soil trophic ecology. In addition, SFWO opens avenues for developing practical ontology-based applications in the field. We also hope this work will foster other ontology development initiatives addressing other aspects of soil ecology.

Box 1. Glossary

Automated reasoning: algorithms and software systems for automating the computation of logical inferences. Typically refers to procedures for deductive reasoning.

Axiom: a logical statement that says what is true in the domain being described. For example, using a subclass axiom, one can state that the class zoo*parasite* is a subclass of the class *parasite.*

Class (concept): a set of individuals that share common characteristics. Class definitions give the properties that the individuals must fulfil to be members of the class.

Domain of discourse: some part of reality, e.g. medicine, ecology, politics. A (domain) ontology is a formal description of the relevant concepts in this domain.

Individual (instance): an actual object from the domain of discourse. Individuals are real-world instances of the concepts defined in the ontology.

Information extraction: the task of automatically transforming natural language text into structured data.

Internationalized Resource Identifier (IRI): a unique sequence of characters that identifies a resource used by web technologies. Ontologies and their elements (classes, properties, individuals…) are identified using IRIs.

Knowledge base: a semantic database resulting from the population of an ontology with a set of individuals. While an ontology describes a domain, a knowledge base (based on an ontology) describes a particular state of affairs.

Property: in an ontology, a property is used to relate two individuals (**object property**) or to relate an individual to a data type value (**datatype property**).

Ontology: a formal model that uses mathematical logic to clarify and explicitly define concepts and their relationships within a domain of interest, e.g. ecology.

Semantic data integration: the process of building a knowledge base by populating an ontology with instance data from multiple heterogeneous sources.

Web Ontology Language: a computational logic-based language for authoring, publishing and sharing ontology on the Web.

Box 2. Abbreviations

BFO	Basic Formal Ontology
CARO	Common Anatomy Reference Ontology
IE	Information Extraction
IRI	Internationalized Resource Identifier

NCBI	National Center for Biotechnology Information
OWL	Web Ontology Language
OBO	Open Biomedical Ontologies
PCO	Population and Community Ontology
RO	Relations Ontology
SFWO	Soil Food Web Ontology
T-SITA	Thesaurus for Soil Invertebrate Trait-based Approaches

Acknowledgements

The research was funded by the French Agence Nationale de la Recherche (ANR) through the GlobNet (ANR-16-CE02-0009) project, from the two 'Investissement d'Avenir' grants (MIAI@Grenoble Alpes: ANR-19-P3IA-0003, OSUG@2020: ANR-10-LAB-56), and the METRO Grenoble Alpes, the Departement Isere and the Office Français de la Biodiversité to Orchamp. Ideas also matured thanks to working group meetings of the EUdaphobase COST Action (CA18237) and sDiv funded sOilFauna project (grant SFW9.02). A.P. acknowledges support of the DFG Emmy Noether program (Project number 493345801) and of iDiv (DFG–FZT 118, 202548816). C.A.M.M. acknowledges the DFG grant MO 412/54-2, project number 326061700, supporting the Specialised Information Service Biodiversity Research (BIOfid, www.biofid.de).

References

[1] Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. Nature, 515(7528), 505-511.

[2] Potapov, A. M., Beaulieu, F., Birkhofer, K., Bluhm, S. L., Degtyarev, M. I., Devetter, M., ... & Scheu, S. (2022). Feeding habits and multifunctional classification of soil-associated consumers from protists to vertebrates. *Biological Reviews*.

[3] Hedde, M., Blight, O., Briones, M. J., Bonfanti, J., Brauman, A., Brondani, M., ... & Capowiez, Y. (2022). A common framework for developing robust soil fauna classifications. *Geoderma*, *426*, 116073.

[4] Potapov, A. M. (2022). Multifunctionality of belowground food webs: resource, size and spatial energy channels. *Biological Reviews*.

[5] Gauzens, B., Legendre, S., Lazzaro, X., & Lacroix, G. (2013). Food-web aggregation, methodological and functional issues. *Oikos*, *122*(11), 1606-1615.

[6] Pey, B., Laporte, M. A., Nahmani, J., Auclerc, A., Capowiez, Y., Caro, G., ... & Hedde, M. (2014). A thesaurus for soil invertebrate trait-based approaches. *PLoS One*, *9*(10), e108985.

[7] Hunt, H. W., Coleman, D. C., Ingham, E. R., Ingham, R. E., Elliott, E. T., Moore, J. C., ... & Morley, C. R. (1987). The detrital food web in a shortgrass prairie. *Biology and Fertility of Soils*, *3*, 57-68.

[8] Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological informatics*, *2*(3), 279-296.

[9] Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., ... & Wooley, J. (2014). Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PloS one*, *9*(3), e89606.

[10] Hoehndorf, R., Schofield, P. N., & Gkoutos, G. V. (2015). The role of ontologies in biological and biomedical research: a functional perspective. *Briefings in bioinformatics*, *16*(6), 1069-1080.

[11] Otte, J. N., Beverley, J., & Ruttenberg, A. (2022). BFO: Basic formal ontology. *Applied ontology*, (Preprint), 1-27.

[12] Schoch, C. (2011). NCBI Taxonomy. 2011 Apr 7 [Updated 2020 Feb 11]. In: Taxonomy Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2011-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK53758/

[13] Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B. D., Clemmensen, K. E., Kauserud, H., ... & Tedersoo, L. (2020). FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Diversity*, *105*(1), 1-16.

[14] Shen, H., Shiratori, Y., Ohta, S., Masuda, Y., Isobe, K., & Senoo, K. (2021). Mitigating N2O emissions from agricultural soils with fungivorous mites. *The ISME Journal*, *15*(8), 2427-2439.

[15] Potapov, A. M., & Tiunov, A. V. (2016). Stable isotope composition of mycophagous collembolans versus mycotrophic plants: do soil invertebrates feed on mycorrhizal fungi?. *Soil Biology and Biochemistry*, *93*, 115-118.

[16] Jassey, V. E., Walcker, R., Kardol, P., Geisen, S., Heger, T., Lamentowicz, M., ... & Lara, E. (2022). Contribution of soil algae to the global carbon cycle. *New Phytologist*, *234*(1), 64-76.

[17] Bolton, J. J. (2016). What is aquatic botany?—And why algae are plants: The importance of non-taxonomic terms for groups of organisms. *Aquatic Botany*, *132*, 1-4.

[18] Moore, J. C., Berlow, E. L., Coleman, D. C., de Ruiter, P. C., Dong, Q., Hastings, A., ... & Wall, D. H. (2004). Detritus, trophic dynamics and biodiversity. *Ecology letters*, *7*(7), 584-600.

[19] Swift, M. J., Heal, O. W., Anderson, J. M., & Anderson, J. M. (1979). *Decomposition in terrestrial ecosystems* (Vol. 5). Univ of California Press.

[20] Xiang, Z., Mungall, C., Ruttenberg, A., & He, Y. (2011, July). Ontobee: A linked data server and browser for ontology terms. In *ICBO*.

[21] Doran, P., Tamma, V., & Iannone, L. (2007, November). Ontology module extraction for ontology reuse: an ontology engineering perspective. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 61-70).

[22] Matentzoglu, N., Goutte-Gattat, D., Tan, S. Z. K., Balhoff, J. P., Carbon, S., Caron, A. R., ... & Osumi-Sutherland, D. (2022). Ontology Development Kit: a toolkit for building, maintaining, and standardising biomedical ontologies. *arXiv e-prints*, arXiv-2207.

[23] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... & Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, *25*(11), 1251-1255.

[24] Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., ... & Simons, N. K. (2019). Towards an ecological trait-data standard. *Methods in Ecology and Evolution*, *10*(12), 2006-2019.

[25] Joimel, S., Nahmani, J., Hedde, M., Auclerc, A., Léa, B., Bonfanti, J., ... & Benjamin, P. (2021, April). A large database on functional traits for soil ecologists: BETSI. In Global Symposium on Soil Biodiversity (pp. 523-528).

[26] Rainford, J. L., & Mayhew, P. J. (2015). Diet evolution and clade richness in Hexapoda: a phylogenetic study of higher taxa. The American Naturalist, 186(6), 777-791.

[27] Le Guillarme, N., & Thuiller, W. (2023). A Practical Approach to Constructing a Knowledge Graph for Soil Ecological Research. bioRxiv 2023.03.02.530763; doi: https://doi.org/10.1101/2023.03.02.530763

[28] Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 523-549.

[29] Farrell, M. J., Brierley, L., Willoughby, A., Yates, A., & Mideo, N. (2022). Past and future uses of text mining in ecology and evolution. *Proceedings of the Royal Society B*, *289*(1975), 20212721.

[30] Le Guillarme, N., & Thuiller, W. (2022). TaxoNERD: deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature. *Methods in Ecology and Evolution*, *13*(3), 625-641.

[31] Eisenhauer, N., Bender, S. F., Calderón-Sanou, I., de Vries, F. T., Lembrechts, J. J., Thuiller, W., ... & Potapov, A. (2022). Frontiers in soil ecology—Insights from the World Biodiversity Forum 2022. Journal of Sustainable Agriculture and Environment, 1(4), 245-261.