bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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[10,11,12], Matty P. Berg[3,4], Maria J.I. Briones[5], Irene Calderón-Sanou[1], Karin Hohberg[6], Camille Martinez Almoyna[1], Carlos A. Martínez-Muñoz[7,8], Benjamin Pey[9], David J. Russell[6], Wilfried Thuiller[1]

Affiliations:

1. Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, F-38000 Grenoble, France.

2. Eco&Sols, INRAE – IRD – CIRAD – Institut Agro, Montpellier, France.

3. Section Ecology and Evolution, Amsterdam Institute of Life and Environment, Vrije Universiteit, De Boelelaan 1085, 1081, HV, Amsterdam, the Netherlands.

4. Community and Conservation Ecology Group, Groningen Institute for Evolutionary Life Science, University Groningen, Nijenborgh 7, 9747 AG Groningen, the Netherlands

5. Dept. Ecologia y Biologia Animal, Universidade de Vigo, Vigo, Spain.

6. Soil Zoology Department, Senckenberg - Leibniz Institution for Biodiversity and Earth System Research, Görlitz, Germany.

7. Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt am Main, Germany

8. Zoological Museum, Biodiversity Unit, University of Turku, Turku, Finland

9. LEFE, Univ. Toulouse – CNRS, Toulouse, France.

10. Soil Biodiversity and Functions, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

11. Faculty of Biology, University of Leipzig, Leipzig, Germany

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Animal Ecology, Johann Friedrich Blumenbach Institut of Zoology and Anthropology, University of Göttingen, Göttingen, Germany.

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

Abstract

Although soil ecology has benefited from recent advances in describing soil organism trophic traits, large scale reconstruction of soil food webs is still impeded by (1) the dissemination of most data about trophic interactions and diets into distributed, heterogeneous repositories, (2) no well-established terminology for describing feeding preferences, processes, and resource types, and (3) much heterogeneity in the classification of different soil groups, or absence of such classifications. Soil trophic ecology could therefore benefit from standardisation efforts. 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, feeding preferences and processes, and provides a formalisation of trophic group definitions. The 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 to facilitate literature-based datasets creation. 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 for better use of available information on the feeding habits of soil organisms and sounder 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 harmonisation, data integration

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Soil multitrophic interactions mediate most of the ecological processes that take place in the soil, including carbon sequestration, nutrient cycling, pest and pathogen regulation. These soil ecological processes support plant health and growth, aboveground biodiversity and ultimately ecosystem resilience and stability (Bardgett and Van Der Putten, 2014). Modelling energy and nutrient transfers between soil organisms involved in multitrophic interactions, through accurate reconstructions of soil food webs, would help to understand the relationships between multitrophic assemblages of soil organisms and ecosystem functioning. However, the immense diversity of soil organisms, which come from all major taxonomic groups present in terrestrial ecosystems, vary in body size over six orders of magnitude (from 0.1 µm to tens of centimetres; Bardgett and Van Der Putten, 2014), and exhibit a wide range of (micro)habitats and feeding preferences (Potapov et al., 2022), has long hampered the reconstruction of accurate soil food webs.

To deal with this diversity in forms and functions, soil biologists resort to classifications of belowground biodiversity into guilds of organisms sharing similar niche requirements, or functional groups of organisms acting similarly on a specific ecological function (Hedde et al., 2022). In particular, when studying multitrophic assemblages, a common approach is to group organisms either in feeding guilds of organisms that feed on the same resources, or in trophic groups of organisms that feed on the same resources and have the same consumers (Hedde et al., 2022). Classifying organisms according to a few trophic groups simplifies food web analysis by allowing the construction of intelligible food web models across the whole spectrum of soil organisms (Potapov, 2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Historically, soil organisms classification into trophic groups was impeded by a persistent lack of consistency in concepts and terminology in soil ecology and by the lack of an overarching framework for classifying the whole soil biota based on their feeding preferences (Hedde et al., 2022). In the absence of such a framework, researchers had to resort to user- or clade-specific definitions of trophic groups, which lead to heterogeneities in the resolution of food webs and limited our ability to draw generic conclusions across studies (Gauzens et al., 2013).

Recent research has worked towards standardisation of trait and trophic group definitions across taxonomic groups. Pey et al. (2014) 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. (2022) developed an integrative classification of soil consumers from protists to vertebrates. This classification uses a hybrid taxonomic-ecological approach, e.g. Oribatida microbivores, Nematoda fungivores, and links group definitions to quantified resource preferences (e.g. plants, algae, litter, bacteria, fungi…) and key traits, including body size, vertical stratification and feeding mechanism. This multifunctional 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. machine-interpretable and standardised) representation of the links between soil organisms' feeding preferences, feeding processes, food resources, and trophic group definitions.

Formal representations of the knowledge in a domain of interest, also called *ontologies*, are increasingly used in biological sciences. Examples in ecology include

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the Extensible Observation Ontology (Madin et al., 2007), the Biological Collections Ontology, the Environment Ontology, the Population and Community Ontology (Walls et al., 2014), and the ECOCORE[1] ontology. Ontologies provide (1) standard identifiers for entities (*classes*) and their relations (*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 relations, e.g. textual definitions, and (4) logical *axioms* and definitions that enable specialised softwares (reasoners) to automatically access, interpret and exploit the knowledge encoded in the ontology (Hoehndorf et al., 2015) (**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	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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

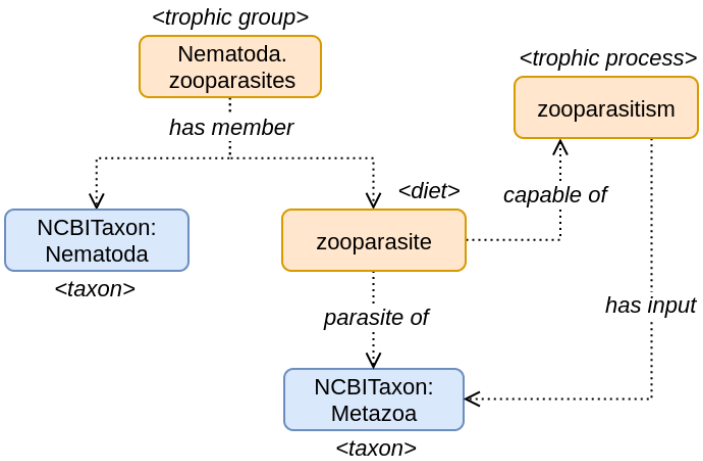

Figure 1. *SFWO provides unique identifiers, synonyms, human (textual) and machine-interpretable (logical) definitions for a collection of concepts related to the diets, food resources, feeding processes, and trophic groups of soil-associated consumers. Logical definitions are constructed by referring to other classes in the ontology, thus creating a logical mapping between the different types of concepts (diets, trophic processes, trophic groups…).*

In this paper, we build on previous efforts to harmonise and standardise terms related to the trophic behaviour of soil-associated consumers (Pey et al., 2014; Potapov et al., 2022) 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 two main objectives. First, it aims to standardise the terminology in use in soil trophic ecology, e.g. by proposing a reference vocabulary to describe information about feeding preferences/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 research traditions, (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. Second, SFWO aims to

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

provide a logical formalisation of the domain of soil trophic ecology, which will facilitate the automated creation of food webs through the use of automated reasoning (**Fig 1.**).

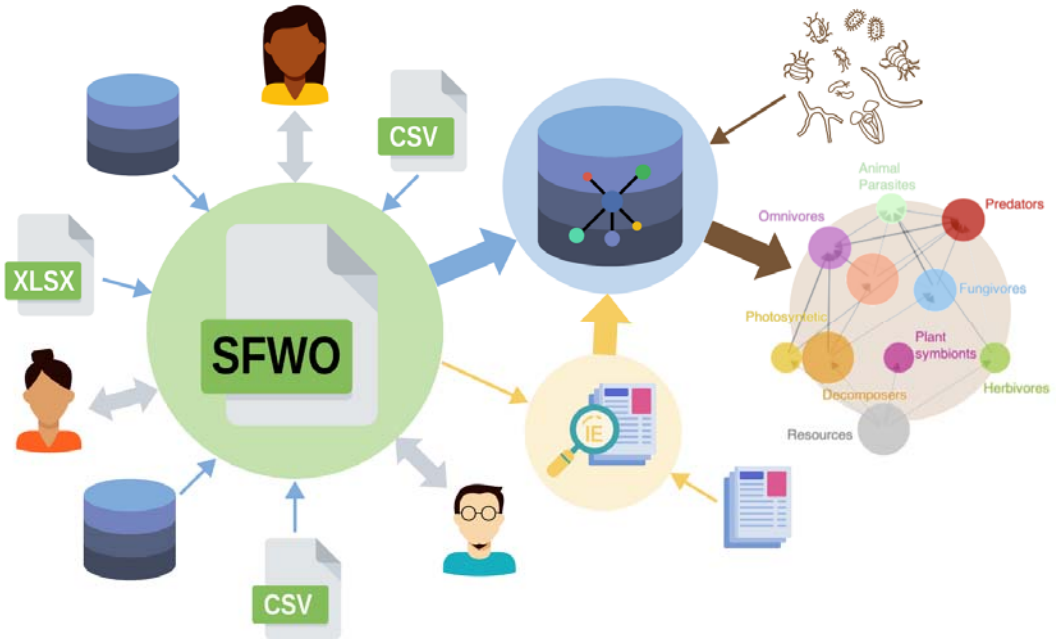

Figure 1. *By providing a standardised terminology and logical formalisation of the domain of soil trophic ecology, SFWO supports knowledge sharing (grey arrows), trophic data integration (blue arrows), extraction of trophic information from literature (yellow arrows), and reconstruction of food webs from soil community data (brown arrows).*

The purpose of this article is to introduce the Soil Food Web Ontology to the soil ecology community, to demonstrate its potential for supporting soil food web research, to encourage soil ecologists to adopt SFWO in the assemblage of their trophic data, and to invite experts in the field to contribute to its development. In the following, we briefly describe the structure and content of the current version of SFWO as well as the ontology development process we adopted. Then we present a case study in which we apply the reasoning capabilities of SFWO to automatically assign a community of soil organisms to their respective trophic groups based on

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

information about their trophic interactions and diets. Finally, we outline the current potential and limits of this ontology for food web research, as well as possible avenues for improvement and broadening of the ontology scope.

Box 1. Glossary

Automated reasoning: algorithms and software systems for automating the computation of logical inferences. Typically refers to procedures for deductive reasoning (Madin et al., 2007).

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. An ontology is a formal description of the relevant concepts in this domain.

Individual (instance): an actual object from the domain. Individuals are the things that the ontology describes or potentially could describe.

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 (Madin et al., 2007).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The Soil Food Web Ontology

SFWO's core structure

SFWO defines a hierarchy of classes (concepts) and properties (relations). The class hierarchy in an ontology acts much like a taxonomy, 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 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_0000053 for *acariphage*, a label (i.e., preferred name, in italics throughout the text), and a human-readable textual definition. Whenever possible, the class or property is assigned a set 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:0000053 for *acariphage*. Wherever possible and necessary to allow deductive 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The basic core structure of SFWO's class hierarchy consists of three main branches, namely *organismal entities*, *food resources* and *trophic processes* (**Fig. 2**). These three concepts are direct descendants of classes from the Basic Formal Ontology (BFO), an upper-level ontology providing foundational terms with well-established semantics that facilitate interoperability between disparate domain 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 (Otte et al., 2022).

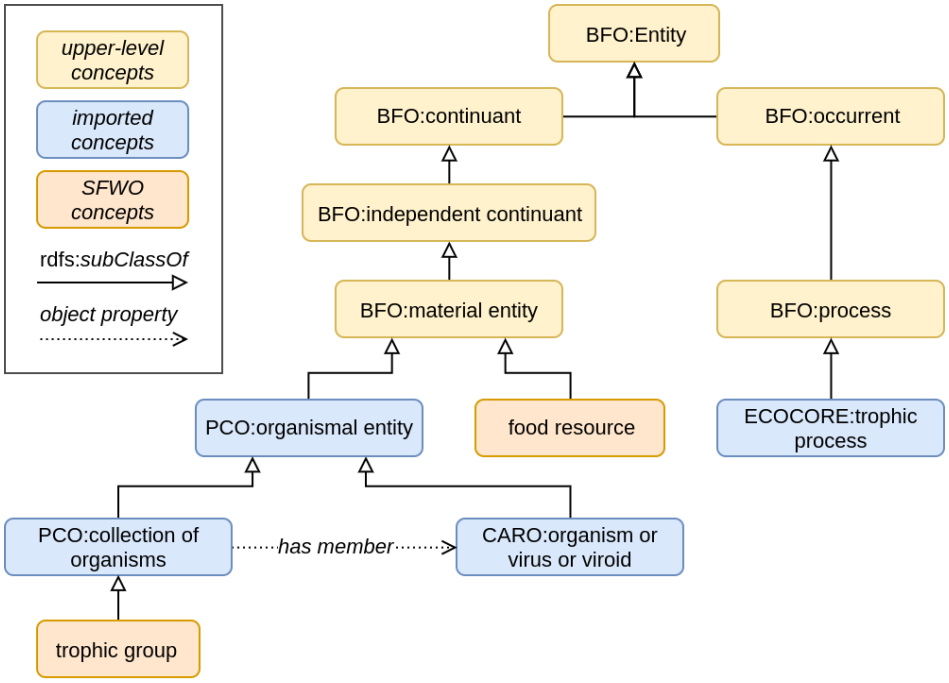

Figure 2. *A high-level view of the three-branch structure of the Soil Food Web Ontology (SFWO). SFWO builds upon the abstract concepts defined in BFO (yellow nodes) for interoperability and reuses a lot of terms from external ontologies (blue nodes). It also defines a number of additional concepts (orange nodes) related to diets, food resources, trophic processes and trophic groups. A solid line between two classes denotes a subClassOf relation, e.g. trophic group is a subclass of collection of organisms. Dotted lines represent semantic relationships between class instances, e.g. a collection of organisms has members that are instances of the organism or virus or viroid class.*

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

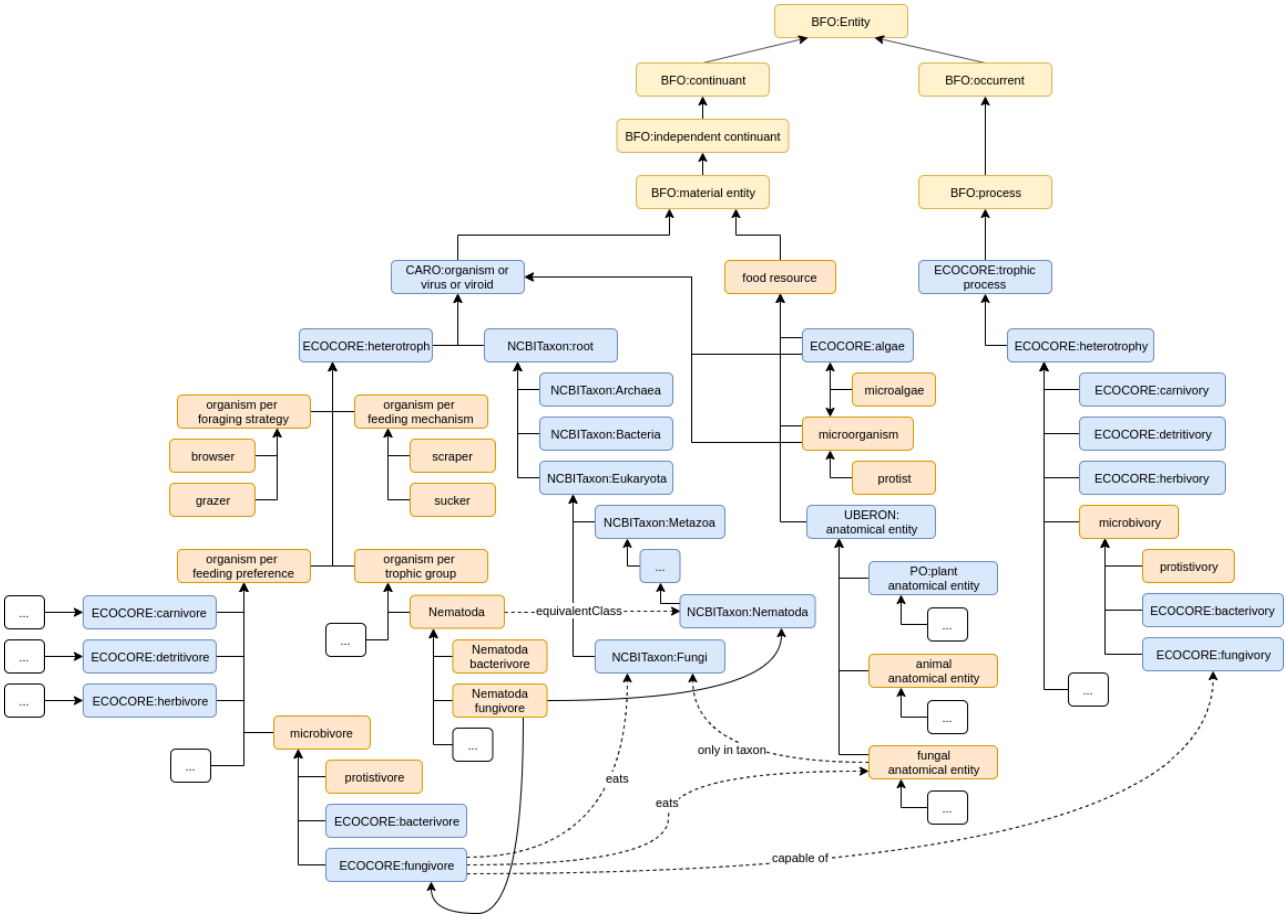

Figure 2. *An excerpt of the Soil Food Web Ontology (SFWO) showing some terms related to feeding preferences, food resources, trophic groups and trophic processes, as well as their relations to upper ontologies. SFWO reuses a lot of terms from external ontologies (blue nodes) and provides a number of new classes (orange nodes). The three main branches (organism, food resource, trophic process) of SFWO are connected by subclassing terms from the Basic Formal Ontology (BFO, yellow nodes). A solid line between two classes denotes a subClassOf relation, e.g. protist is a subclass of microorganism. Dashed lines represent other semantic relationships between classes or class instances, e.g. a fungivorous organism is logically defined as an organism that eats fungi or fungal parts.*

Organisms. 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 part

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of the SFWO ontology, the first one grouping concepts related to taxa, diets and non-taxonomic organisms, the second one providing a hierarchy of trophic groups as collections of organisms sharing the same resources and consumers.

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

All taxonomic classes are imported from the NCBITaxon ontology which is an automatic translation of the NCBI taxonomy database (Schoch, 2011) into OWL. To keep the ontology as compact as possible, SFWO imports a taxonomic class (and all its superclasses corresponding to higher taxonomic ranks) only if this class is used as part of the logical definition of another class. For instance, the taxon *Nematoda* [NCBITaxon:6231] is used in the logical definitions of the diet *nematophage* [SFWO:0000019] and trophic group *Nematoda.zooparasites* [SFWO:0000279].

On the other hand, some organismal entities commonly found in soil food web research cannot be assigned to a taxonomic group. This is the case for example of *algae* [ECOCORE:00000076] which is a non-taxonomic group representing a large polyphyletic group of diverse photosynthetic organisms that excludes vascular plants and mosses and includes cyanobacteria, seaweed, and single-celled microalgae and protists. 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 diet classes.

SFWO reuses many nutrition-related classes from ECOCORE, an ontology of core ecological entities. This includes, terms that refers to an organism's source of carbon/energy/electron, e.g. *heterotroph* [ECOCORE:00000010], *chemotroph*

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[SFWO:0000000], *organotroph* [SFWO:0000001], that are all immediate subclasses of *organism or virus or viroid*. 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. Among heterotrophic organisms, we distinguish between organisms according to their diet (e.g. *nematophage*, *coprophage*, *plant parasite*), feeding mechanism (e.g., *scraper*, *sucker*, *shredder*), and foraging strategy (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* [SFWO:0000019], whose textual 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 imported from the Relations Ontology and used to assert a trophic interaction between two individuals. The logical definition broadly matches the textual definition[2]. Logical definitions of diet classes are key to our

[2] Actually, similarly to textual definitions, logical definitions should preferably adopt the genus-differentia form, such 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 *myrmecophage*

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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).

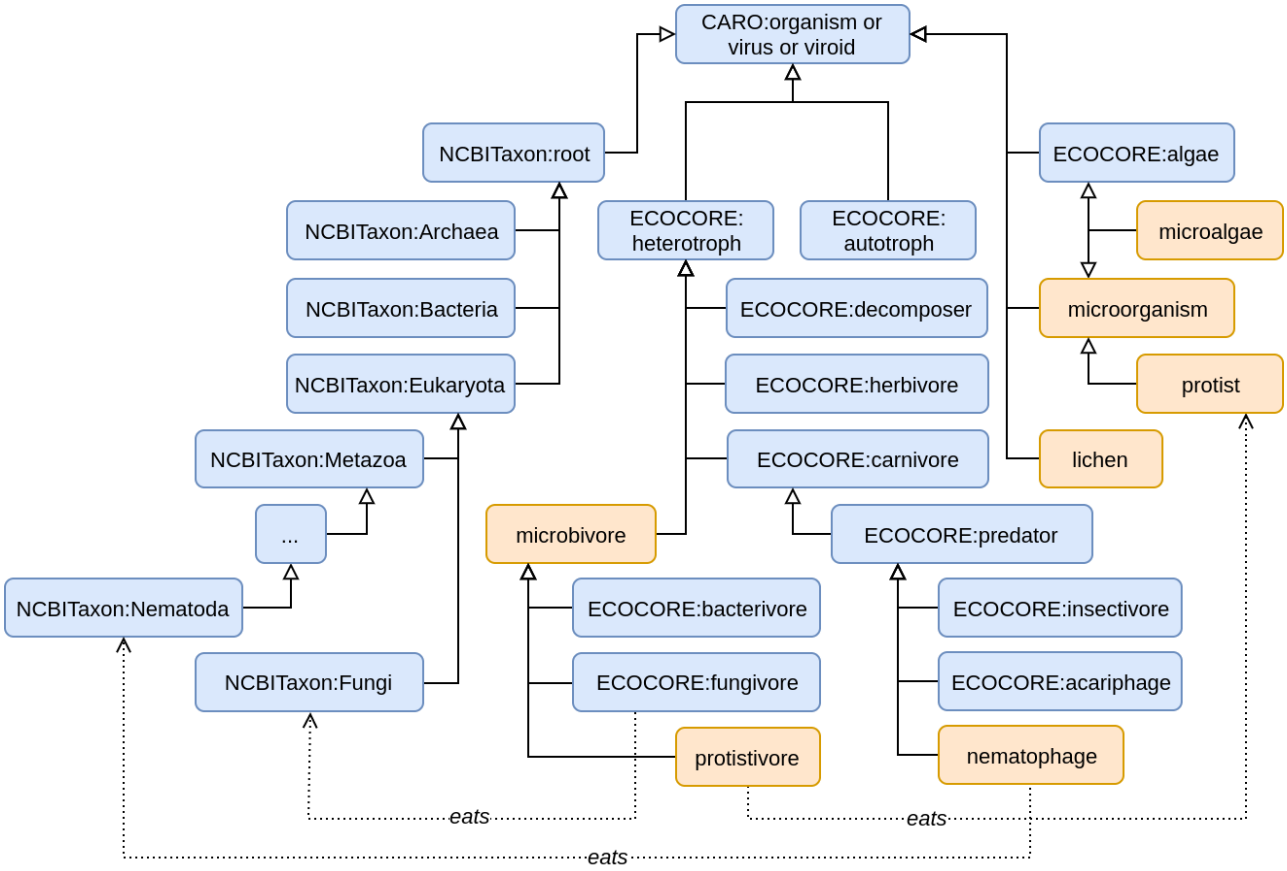

Figure 3. *An excerpt from the organism branch of SFWO. Logical definitions of diet classes create mappings between diets and trophic interactions (consumer-resource relations).*

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 of Hedde et al. (Hedde et al., 2022). In the current version of SFWO, this branch of the ontology expands into a hierarchy of trophic groups for soil-associated protists and animals and fungi.

---

should be equivalentTo(*myrmecophage, insectivore and eats* some *Formicidae*), where *insectivore* is the genus and *eats* some *Formicidae* is the differentia. In SFWO, we ignore this best practice for diet classes as it makes inferring an organism diet from the food resources it consumes more difficult. In a reference ontology unconstrained by reasoning applications, we would stick to the genus-differentia form.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Trophic groups for protists and animals are taken from the taxonomic-ecological classification of soil-associated consumers proposed by Potapov et al. (Potapov et al., 2022 - 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. Examples of nematode trophic groups from Potapov et al. are given in Fig. 3.

Fungal trophic groups are taken from the FungalTraits database (Põlme et al., 2020). Among other functional traits, FungalTraits provides data about the trophic modes (*saprotroph*, *symbiotroph*, *pathotroph*) and feeding guilds (e.g. *litter saprotroph*, *arbuscular mycorrhizal, animal pathogen*) for fungi and fungus-like organisms. Adopting the same hybrid approach as for protist and animal trophic groups, we turned FungalTraits trophic modes and feeding guilds into a hierarchy of fungal trophic groups (see examples in Fig 3.).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

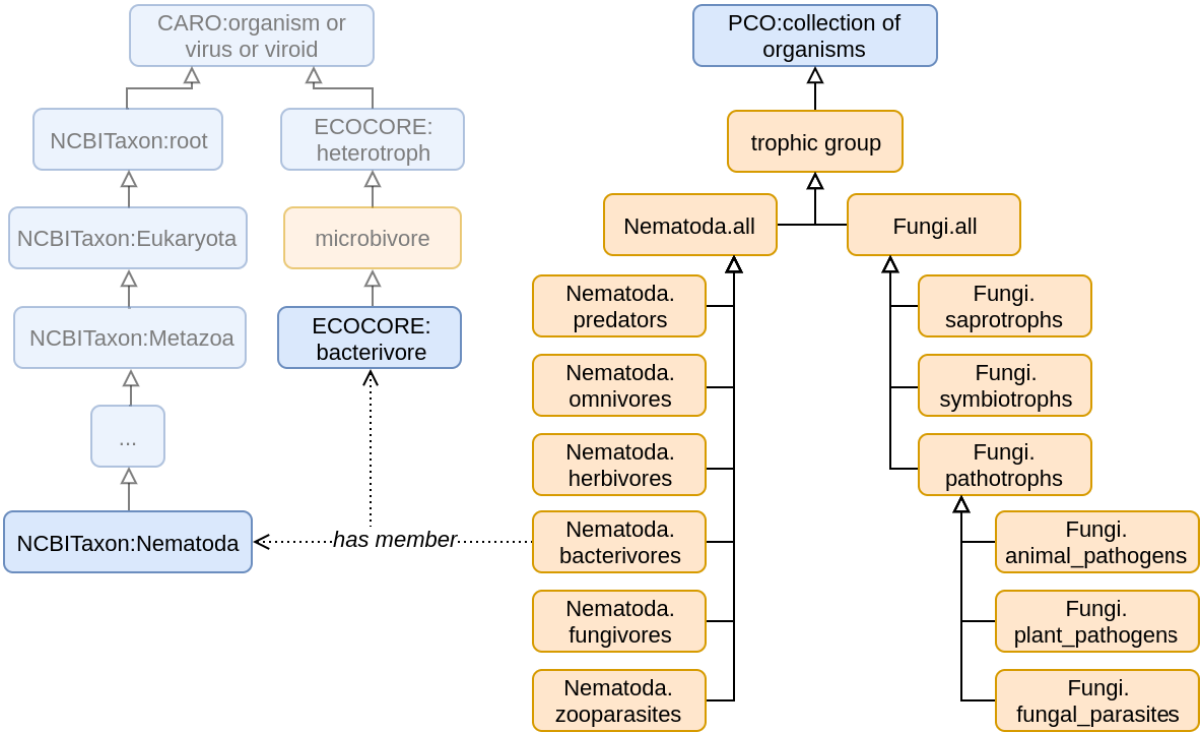

Figure 3. *An excerpt of the hierarchy of trophic groups in SFWO. Trophic group definitions are taken from (Potapov et al., 2022) and (Põlme et al., 2020). The complete hierarchy consists of 165 trophic group classes for 42 taxa representative of soil biota, from protists to fungi. Logical definitions of trophic group classes create a mapping with taxonomic classes and organism traits, e.g. diets.*

SFWO also provides logical definitions for most of these trophic group classes. These 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* states that members of this trophic group are necessarily bacterivorous nematodes. 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] <u>and</u> *has member* [RO:0002351] <u>only</u>

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	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 logical formalisation of the trophic classification of nematodes (Nematoda) in SFWO, according to the taxonomic-ecological classification of Potapov et al. (Potapov et al., 2022). The classification is hierarchical, and the root class for a given taxon (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". Loosely speaking, 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) anatomical structures, 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, 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 disseminated across different branches of the ontology. However, we decided to create a convenience class *food resource* to gather under a common root concept all the entities that are commonly considered as trophic resources in the literature or for food-web reconstruction. This could prove useful for text mining

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

applications (e.g. extracting consumer-resource relations from the literature) 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 the ontology.

Trophic processes. Mirroring the hierarchy of classes related to feeding preferences (diets), SFWO provides a hierarchy of terms referring to trophic processes under the root term *trophic process* [ECOCORE:00000008] (see **Fig. 2**). For sake of completeness, all the nutrition-related classes have their processual counterparts, e.g. coprophagy [SFWO:0000043] is the trophic process involving a coprophagous consumer (*coprophage* [ECOCORE:00000176]) feeding on faeces. 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 *animal parasite* 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 mapping between trophic preferences 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 used most often as part of the diet

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 a trophic interaction but a symbiotic interaction in RO. 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.

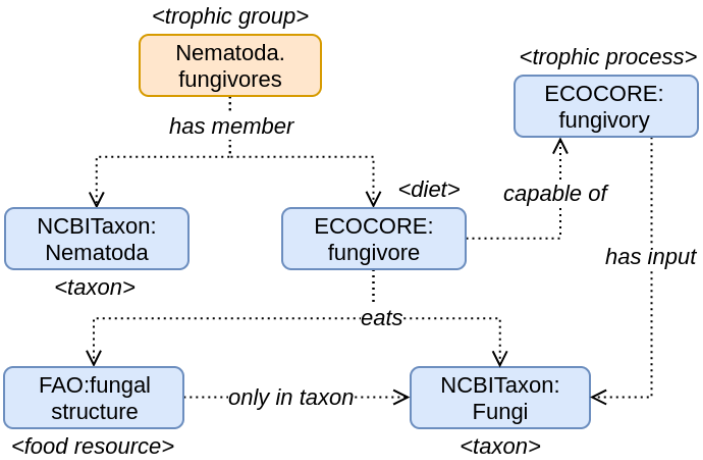

Figure 4.

Reasoning in SFWO

Logical axiomatization of classes creates a mapping between taxonomic entities, trophic groups, diets, food resources and trophic processes (see Fig. 4). Creating axiomatized class definitions allow a reasoner to automatically classify individuals using deductive reasoning. In particular, this allows for three types of deduction:

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In addition, SFWO includes a number of general class axioms, i.e. logical axioms with anonymous classes on the left-hand side. For instance,

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

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

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

An example of deductive inference is given in Fig. 5. In this example, the fact that *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 deduce, using SFWO's logical axioms, the following facts:

1. i_1 is a *Nematoda* (because *A. avenae* is a subclass of *Aphelenchus*, 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* some *(Fungi* or *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*).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

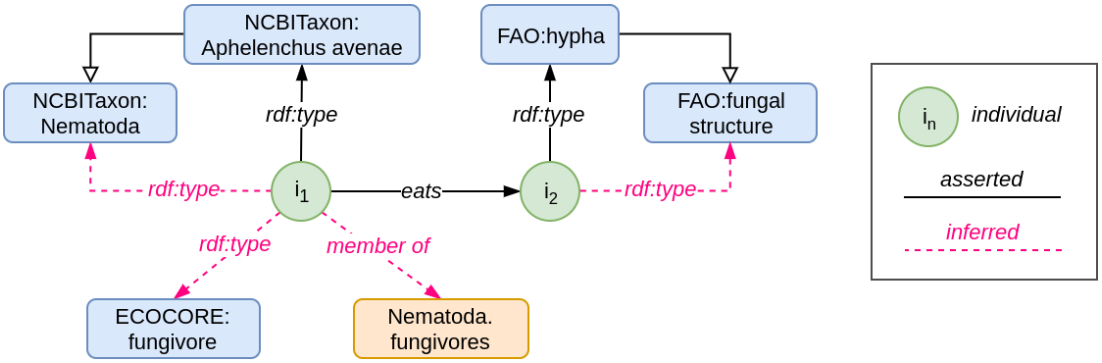

Figure 5. *A reasoner is able to infer logical consequences from asserted facts and axioms. Here, the diet and trophic group of A. avenae are deduced from the fact that this species feeds on hyphae.*

Terminological issues

The development of SFWO has highlighted a number of terminological issues currently present in the research field that were, at least partly, addressed by us (see terminological discussions on GitHub [link?]). Below we describe a few of these issues as examples of the challenges of harmonising the terminology in a scientific domain and modelling these concepts as part of a formal representation of the domain. In our decisions we followed 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 trophic groups in soil food webs. For example, organisms feeding on fungi are named *fungivores* (e.g. Shen et al. 2021), *mycophages*, or *mycotrophs* (e.g. Potapov and Tiunov, 2016) in different literature sources. Following the literature and expert exchange, we uniformly used *-vores* and *-phages* as synonyms for organisms that feed on specific *food resource* by consuming the whole/part of it (*herbivores* = *phytophages*, *bacterivores* = *bacteriophages*); *-trophy* is used as a broader term depicting any type

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of energy/nutrient acquisition from specific *food resource*, including e.g. symbiotic exchange (mycorrhizal fungi, *symbiotrophs* {XXX?}) or extracellular degradation and intake of organic matter (microbial decomposition process, implemented by *saprotrophs* {XXX?}; GitHub [link1, link2?]).

Another terminological issue represented non-taxonomic groups such as *algae*. Because *algae* is a polyphyletic group of organisms, it is challenging to formally align *algivory* {XXX?} with *herbivory* {XXX?} (implying feeding on living plant tissues) and *microbivory* {XXX?} (implying feeding on bacteria, fungi, and/or protists – the latter comprising many soil *microalgae* {XXX?}; Jassey et al., 2022). In SFWO, *algae* are considered as a subclass of *microorganisms* and *algivory* is a subclass of *microbivory*, because algae in soil are unicellular (i.e. *microalgae*) and are consumed mainly by bacterial, fungal, and protist feeders, rather than 'true' herbivores (i.e. root-feeding nematodes, curculionid larvae; Potapov et al., 2022). However, this decision is suboptimal as we probably need an easy term to depict trophic processes related to primary production (Bolton, 2016; GitHub [link?])).

A broad discussion was about the scope and classification of *detritus*. Moore et al. (2004) define *detritus* based on Swift et al. (1979) 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. extra-cellular polymers, nectar, root exudates and leachates, dissolved organic matter, extra-cellular matrix, mucilage)*." This broad definition includes nectar and root exudates, and as such we must classify rhizosphere microorganisms, nectar feeders (e.g. bees), and any body surface-dwelling symbionts as *detritivores* or *saprotrophs*. Nectar and exudates are

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

connected to surfaces of a living organism, and at present in SFWO we classify *nectarivores* as a subclass of *herbivores* (i.e. feeding on living plant matter), because the opposite would have a consequence of classifying numerous nectar-feeding insects as *detritivores* if the ontology will be expanded to aboveground food webs (GitHub [link1? Link2? For detritus classification]). *Exudativores* are at present also included in *herbivores* – an example can be rhizosphere bacteria, which are abundant within the first 4 mm around the root… (Kuzyakov and Razavi, 2019).

SFWO's development process

The ontology design process started with a small team of experts in soil invertebrate ecology that provided a list of 100+ terms related to primary nutritional groups, diets, and food resources of soil-associated consumers, as well as their definitions and relationships between diets and food resources, e.g. *entomophage* → *Insecta*, *saproxylophage* → *dead wood*. A subset of these core terms was manually aligned with classes from existing ontologies using Ontobee's search engine (Xiang et al., 2011). For instance, the term bacterivore matches the class *bacterivore* [ECOCORE:00000098] defined in ECOCORE. These classes were used as "seeds" for module extraction. Module extraction is a popular strategy for ontology reuse that avoids the overheads involved in importing complete ontologies (Doran et al., 2007). Modules were merged to form the backbone of the Soil Food Web Ontology. We then created new classes for terms that could not be imported from existing ontologies using Protégé[3], a free and open-source ontology editor. The ontology

[3] https://protege.stanford.edu/

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

development workflow is managed and automated using the Ontology Development Kit (Matentzoglu et al., 2022).

As far as possible, we strived to make SFWO conform to the shared principles for ontology development prescribed by the Open Biomedical Ontologies (OBO) Foundry (Smith et al., 2007), including the use of the Basic Formal Ontology as an upper level ontology, and the mandatory reuse 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. Although interoperability with existing ontologies is an important concern, we sometimes had to modify an imported class to better reflect the correct use of the concept in soil trophic ecology. While some "additive" operations, e.g. adding a new subclass, a new synonym…, have no impact on the intended meaning of the imported class, other operations may change its logical definition, e.g. removing a superclass, changing an equivalent class axiom. As a consequence, SFWO cannot be merged back with the source ontology, which prevents reuse of both ontologies in a different context. As an ontology for soil trophic ecology with a strong focus on applications, e.g. soil food web reconstruction, we decided to sacrifice reuse as part of the OBO ecosystem for sake of representational fidelity and reasoning tractability.

SFWO is continuously updated, improved and enriched thanks to users' 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

using the issue tracker of the project's GitHub repository[4]. Requests for changes/modifications are regularly submitted to the steering committee for validation. Once approved, changes are applied using the above methodology (extraction of modules or creation of new classes, manual editing to add synonyms or to modify textual or logical definitions...) and a revised version is published.

Downloads

SFWO's latest release version is available for download on the project's website [CITE]. The ontology is available in OWL format and consists of a single file (sfwo.owl). The ontology can also be browsed online on AgroPortal[5]. The ontology is under version control in a GitHub repository which is also used to track changes and issues and manage releases. The version of the ontology described in this manuscript is available from http://github.com/nleguillarme/soil_food_web_ontology/releases/download/v2023-XX-XX/sfwo.owl.

Case study

The study of soil multi-trophic diversity and food web structure can help unveil the intricate linkages between soil biodiversity and ecosystem functioning, and the way these are impacted by environmental changes, e.g. land-use transformation (Barnes et al., 2014), agricultural intensification (Bloor et al., 2021), insect outbreaks (Calderón-Sanou et al., 2021)... Macroecological studies based on the construction of comprehensive food webs for soil communities have benefited from a number of advances: environmental DNA and metabarcoding technology have emerged as a

[4] http://github.com/nleguillarme/soil_food_web_ontology
[5] http://agroportal.lirmm.fr/ontologies/SFWO

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

promising tool for soil biodiversity monitoring ; the increasing availability of open databases on biotic interactions and functional guilds facilitate the assignment of trophic groups. Yet, trophic group assignment remains primarily a manual process, heavily influenced by the taxonomic resolution of community composition data and the information available on the diets or trophic position of identified taxa. This impacts the resolution of the reconstructed food webs, which in turn impacts the ability of the food web model to reliably and meaningfully capture specific effects of disturbance on some taxonomic groups (Calderón Sanou et al., 2021). This case study aims to demonstrate the potential of SFWO in automating the trophic group assignment process, given a sufficiently large database of trophic interactions.

Applications of the Soil Food Web Ontology

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. More specifically, SFWO was 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

As stressed by Schneider et al. (2019), the lack of standardisation for trait data, e.g. diet data, hampers the efficient reuse and synthesis of published datasets. A major challenge in trait data standardisation is the lack of widely accepted and unambiguous trait definitions. The T-SITA thesaurus (Pey et al., 2014) is an effort to provide such definitions for soil invertebrates. Out of many non-trophic traits, T-SITA

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

defines a hierarchy of 20+ diet-related terms. SFWO goes one step further. With an ever-increasing vocabulary of XXX terms for describing trophic interactions, diets, food resources and trophic processes encountered, SFWO has the potential to become the standard vocabulary for describing, storing and sharing soil trophic data, thus facilitating their reuse for synthesis and integration into larger databases. In addition to human-readable definitions, and contrary to T-SITA, SFWO also includes machine-interpretable logical axioms that allow for inference of additional knowledge using deductive reasoning.

Trophic data integration and knowledge base construction

SFWO was designed to ease semantic data integration, i.e. the process of combining data from multiple sources into a single knowledge base using ontologies. Ecology entered the Big Data era for a few years now, yet large-scale studying of ecological systems is still hampered by the fact that biodiversity data are highly heterogeneous and widely distributed, although increasingly available. Semantic data integration, has the potential to facilitate large synthetic ecological studies by (1) providing a unified access to distributed data sources, (2) making heterogeneous data interoperable using a unified representation, and (3) providing data with semantic context ("meaning") to support its interpretation. SFWO is an important contribution for trophic data integration, as it provides the semantic foundations for representing knowledge in the domain of soil trophic ecology, and building a unified knowledge base of trophic data for the whole soil biota. This requires integrating feeding-related trait data from distributed sources, e.g. Global Biotic Interactions, FungalTraits, BETSI…, into SFWO using some ontology-based data integration

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pipeline, such as the ontology-data-pipeline[6] or inteGraph[7]. Such a knowledge base will be a valuable resource for macroecological studies of soil ecosystems, soil food web reconstruction, trophic information extraction, etc.

Soil food web reconstruction

Several aspects of SFWO can support the reconstruction of soil food webs from community composition data. Firstly, by facilitating the integration of trophic data from different sources, SFWO contributes to fill the "Eltonian shortfall" in soil ecology, i.e. a lack of knowledge about inter- and intra-specific relationships within soil communities. Secondly, with over XXX axiomatized trophic group classes covering a broad spectrum of soil taxonomic groups, SFWO provides a common, logically-sound framework for soil-associated consumers classification. This framework solves issues related to inconsistent definitions of trophic groups across studies and the heterogeneity in food web aggregation levels that they entail, thus making food webs reconstructed in different research contexts reusable, comparable and combinable as part of 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.

Trophic information extraction

A number of the above applications of SFWO depend on our ability to collect large species-level 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,

[6] https://github.com/biocodellc/ontology-data-pipeline
[7] https://github.com/nleguillarme/inteGraph

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

screening, data extraction, and harmonisation of entities, such as species or trait names (Farrell et al., 2022). An alternative is to use information extraction approaches to automatically turn unstructured text into structured data. With its terminology containing more than XXX terms for over YYY concepts related to soil trophic ecology, SFWO is a valuable lexical resource for 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. SFWO is therefore particularly suitable to detect mentions of trophic groups, trophic processes, food resources and diets in text. In combination with a taxonomic entity recognition system (Le Guillarme and Thuiller, 2022), SFWO could support the automatic creation, curation and updating of literature-based datasets.

Beyond the soil : towards a Food Web Ontology

Conclusion

With XXX concepts, YYY synonyms, ZZZ axiomatized diet classes and a full axiomatization of the multifunctional classification of Potapov et al. (2022), we believe that SFWO offers an immediately useful resource to support soil ecology researchers in the collection, annotation, interpretation and exploitation of trophic data. In the near future, we aim to increase the expressiveness of SFWO, notably by introducing mechanisms to represent uncertainty in diet assignment and trophic position, to quantify resource preferences, … also add bacteria classification, other traits (all trophic groups from Potapov et al. cannot be axiomatized).

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A good ontology is consensual in nature, which means it should capture domain knowledge in a way that is accepted by the community. This is also key to a widespread adoption of a semantic model to the point it becomes a standard. To help build and maintain consensus over the years, SFWO's development has been conceived as an ongoing and collaborative process. Therefore, on behalf of the steering committee, we invite domain experts to contribute their knowledge in the development of the ontology through the project website. In addition, we encourage any person interested in using SFWO and who would be unfamiliar with this kind of semantic technologies to contact us, also through the project website.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

References

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

Barnes, A. D., Jochum, M., Mumme, S., Haneda, N. F., Farajallah, A., Widarto, T. H., & Brose, U. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nature communications*, *5*(1), 1-7.

Bloor, J. M., Si-Moussi, S., Taberlet, P., Carrère, P., & Hedde, M. (2021). Analysis of complex trophic networks reveals the signature of land-use intensification on soil communities in agroecosystems. *Scientific Reports*, *11*(1), 1-10.

Briones, M. J. I. (2014). Soil fauna and soil functions: a jigsaw puzzle. *Frontiers in Environmental Science*, *2*, 7.

Calderón-Sanou, I., Münkemüller, T., Zinger, L., Schimann, H., Yoccoz, N. G., Gielly, L., ... & Thuiller, W. (2021). Cascading effects of moth outbreaks on subarctic soil food webs. *Scientific reports*, *11*(1), 1-12.

Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and Biogeography*, *19*(3), 287-302.

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).

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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gardner, S. P. (2005). Ontologies and semantic data integration. *Drug discovery today*, *10*(14), 1001-1007.

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

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.

Jackson, R. C., Balhoff, J. P., Douglass, E., Harris, N. L., Mungall, C. J., & Overton, J. A. (2019). ROBOT: a tool for automating ontology workflows. *BMC bioinformatics*, *20*(1), 1-10.

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.

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.

Madin, J. S., Bowers, S., Schildhauer, M. P., & Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends in ecology & evolution*, *23*(3), 159-168.

Majdi, N., Hette-Tronquart, N., Auclair, E., Bec, A., Chouvelon, T., Cognie, B., ... & Perga, M. E. (2018). There's no harm in having too much: A comprehensive toolbox of methods in trophic ecology. *Food webs*, *17*, e00100.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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.

Potapov, A. M., Pollierer, M. M., Salmon, S., Sustr, V., & Chen, T. W. (2020). Multidimensional trophic niche approach: gut content, digestive enzymes, fatty acids and stable isotopes in Collembola. *bioRxiv*, 1-47.

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.

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

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*.

Scheu, S. (2002). The soil food web: structure and perspectives. *European journal of soil biology*, *38*(1), 11-20.

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.

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/

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.03.526812; this version posted February 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

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.

Wimalasuriya, D. C., & Dou, D. (2010). Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science*, *36*(3), 306-323.

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