

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

3 **List of authors:** Nicolas Le Guillarme¹, Mickael Hedde², Anton M. Potapov^{3,4,5}, Matty P. Berg^{6,7}, Maria
4 J.I. Briones⁸, Irene Calderón-Sanou^{1,3}, Florine Degruene^{2,9}, Karin Hohberg¹⁰, Camille Martinez-
5 Almoyna¹, Carlos A. Martínez-Muñoz^{11,12}, Benjamin Pey¹³, David J. Russell¹⁰, Wilfried Thuiller¹

7 **Affiliations:**

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

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

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

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

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

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

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

18 8. Departamento de Ecología y Biología Animal, Universidade de Vigo, Vigo, Spain.

19 9. CIRAD, UMR Eco&Sols, Montpellier, France

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

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

24 12. Zoological Museum, Biodiversity Unit, University of Turku, Turku, Finland

25 13. LEFE, Univ. Toulouse – CNRS, Toulouse, France.

26

27

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

29 **Abstract**

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

47

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

49 **Introduction**

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

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

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

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

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

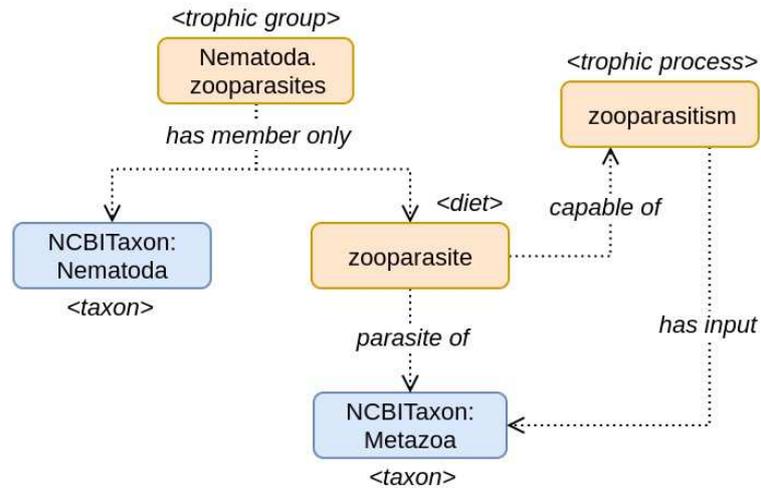
87 interpretable) representation of the links between soil organisms' diets (food types), feeding
 88 processes, food resources, and trophic groups.

89 Formal representations of the knowledge in a domain of interest, also called *ontologies*, are
 90 increasingly used in biological sciences. Examples include the Extensible Observation Ontology [8],
 91 the Biological Collections Ontology, the Environment Ontology, the Population and Community
 92 Ontology [9], and the ECOCORE¹ ontology of core ecological entities. Ontologies provide (1) unique
 93 identifiers for the concepts (*classes*) and their relationships (*properties*) within a *domain of discourse*,
 94 (2) vocabularies of the terms used within this domain, (3) metadata describing the intended meaning
 95 of the classes and properties, e.g. textual definitions, and (4) logical *axioms* and definitions that
 96 enable specialised softwares (*reasoners*) to automatically access and interpret the knowledge
 97 encoded in the ontology [10] (Figure 1). These features make ontologies a powerful tool that facilitates
 98 the harmonisation, interpretation, sharing and reuse of domain knowledge, eases the process of
 99 integrating heterogeneous data from multiple sources (*semantic data integration*), enables inference
 100 of new knowledge using *automated reasoning*, and supports literature-based database creation and
 101 curation using *information extraction*.

102

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

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



103

104

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

109

110

111 In this paper, we build on previous efforts to harmonise and standardise terms related to the trophic

112 behaviour of soil-associated consumer [2,6] to develop the Soil Food Web Ontology (SFWO), a formal

113 representation of domain knowledge in the field of soil trophic ecology. The development of SFWO

114 pursues several objectives (Figure 2). First, it aims to standardise the terminology in use in soil trophic

115 ecology, e.g. by proposing a reference vocabulary to describe information about diet and feeding

116 processes and harmonising the definitions of trophic groups across taxonomic groups. This will allow

117 to: (1) facilitate knowledge sharing between experts on different taxonomic groups and from different

118 disciplines, (2) facilitate the integration of heterogeneous trophic datasets, (3) support information

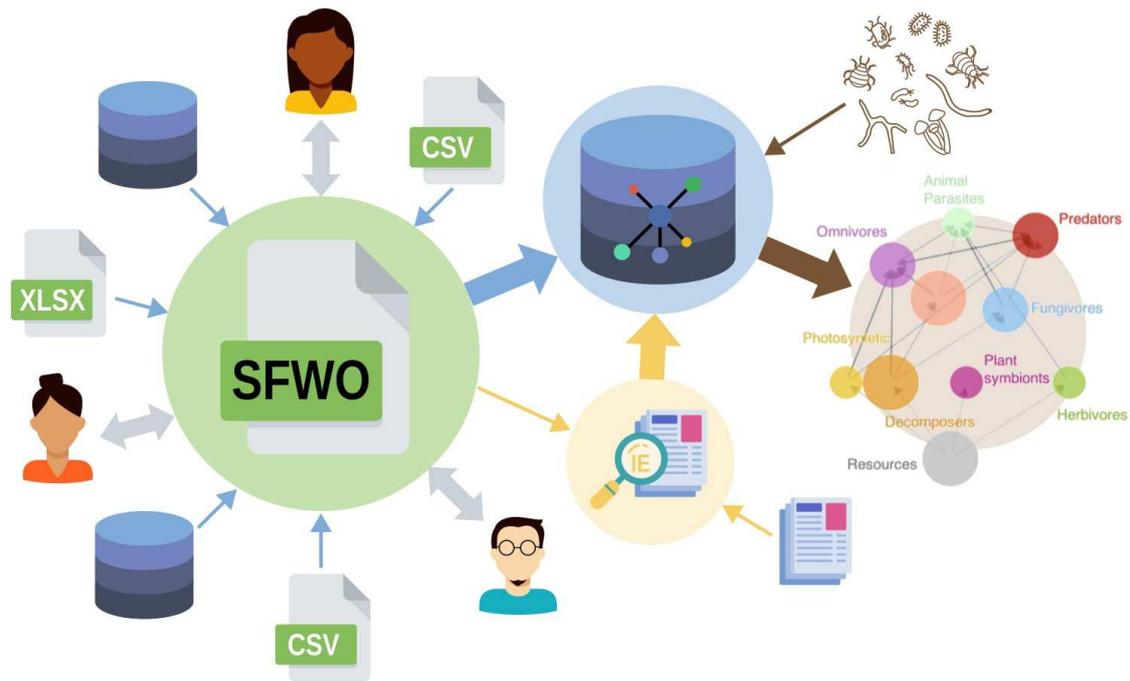
119 extraction applications by providing a comprehensive list of terms related to soil trophic ecology, and

120 ultimately to address more complex food-web related research questions in large synthetic studies.

121 Secondly, SFWO aims to provide a logical formalisation of the domain of soil trophic ecology, which

122 will facilitate the creation of food webs through the use of automated reasoning.

123



124

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

129

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

137 Materials and Methods

138 The Soil Food Web Ontology

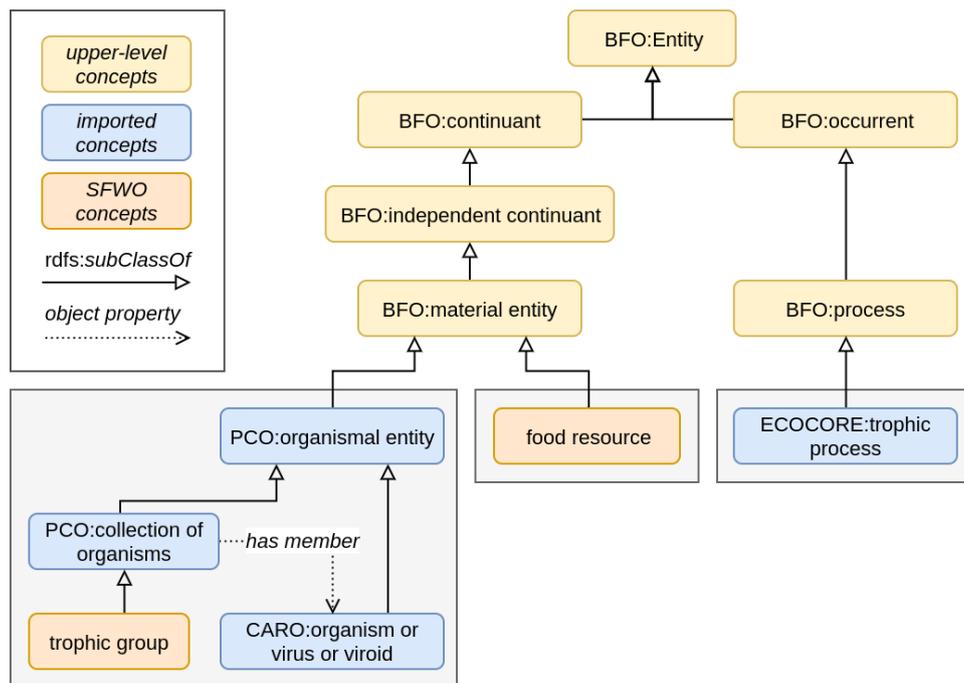
139 SFWO defines a hierarchy of *classes* (concepts) and *properties* (relations). The class hierarchy in an
140 ontology acts much like a taxonomy of concepts, with children (or subclasses) inheriting the properties
141 of their parents (or superclasses). A child class is a specialisation of a parent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

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

157

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

165



166

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

174

175

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

185

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

190 *Taxon*. All taxonomic classes are imported from the NCBITaxon ontology² which is an automatic
191 translation of the NCBI taxonomy database [12] into OWL and is currently the only taxonomy available
192 in ontological form. To remain as compact as possible, SFWO imports a taxonomic class (and any
193 superclasses in its taxonomic lineage) only if that class is used as part of the logical definition of
194 another class. For instance, the taxonomic class *Nematoda* [[NCBITaxon:6231](#)] is used in the logical
195 definitions of the diet *nematophage* [[SFWO:0000019](#)] and trophic group *Nematoda.zooparasites*
196 [[SFWO:0000279](#)].

197

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

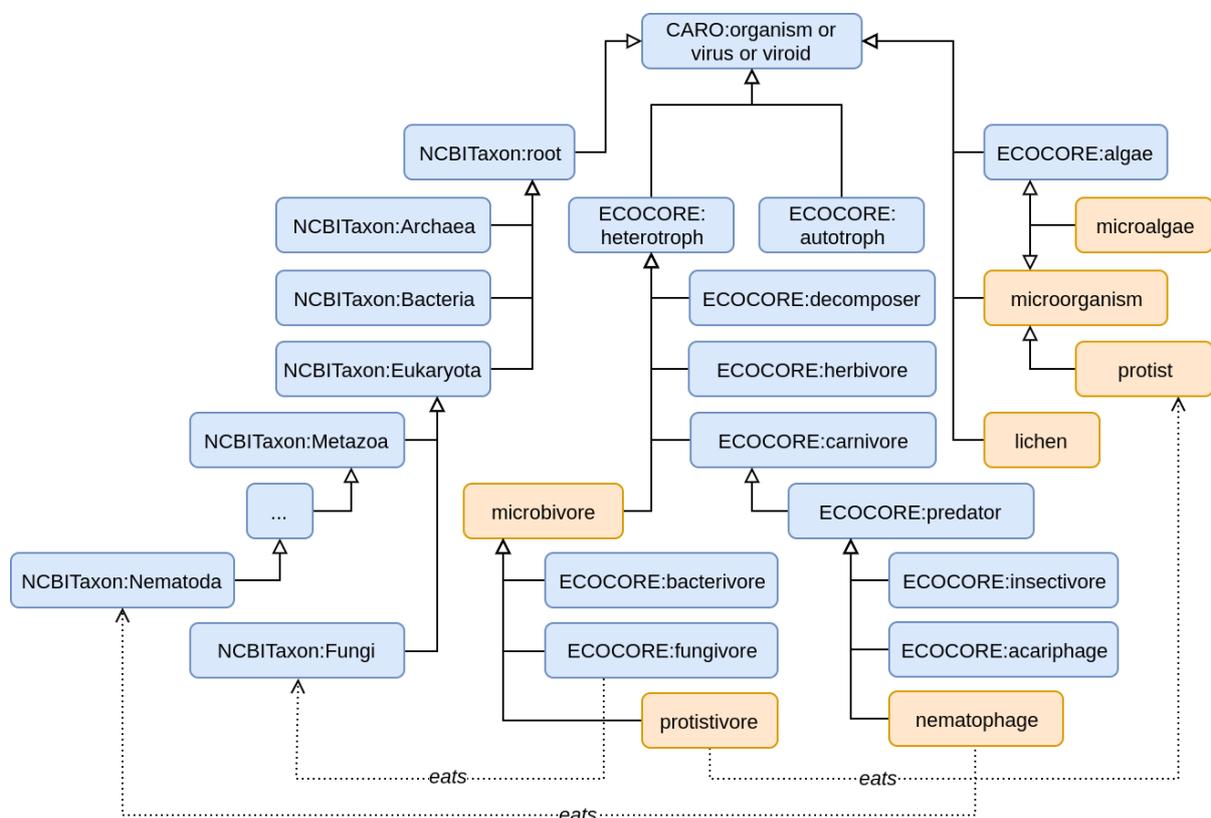
205

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

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

221 $\text{equivalentTo}(\text{nematophage}, \text{eats some Nematoda}),$

222 which means that any individual involved in an eats [RO:0002470] relationship with at least one
 223 individual of class *Nematoda* [NCBITaxon:6231] is a *nematophage*. Here, eats is an object property (=
 224 relationship) imported from the Relations Ontology³ (RO) and used to assert a trophic interaction
 225 between two individuals. The logical definition of *nematophage* broadly matches the textual definition⁴.
 226 Logical definitions of diet classes are key to our ability to use ontological reasoning to automatically
 227 infer an organism diet(s) and trophic group(s) from its trophic interactions (see subsection on
 228 reasoning in SFWO).



229

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

232

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

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

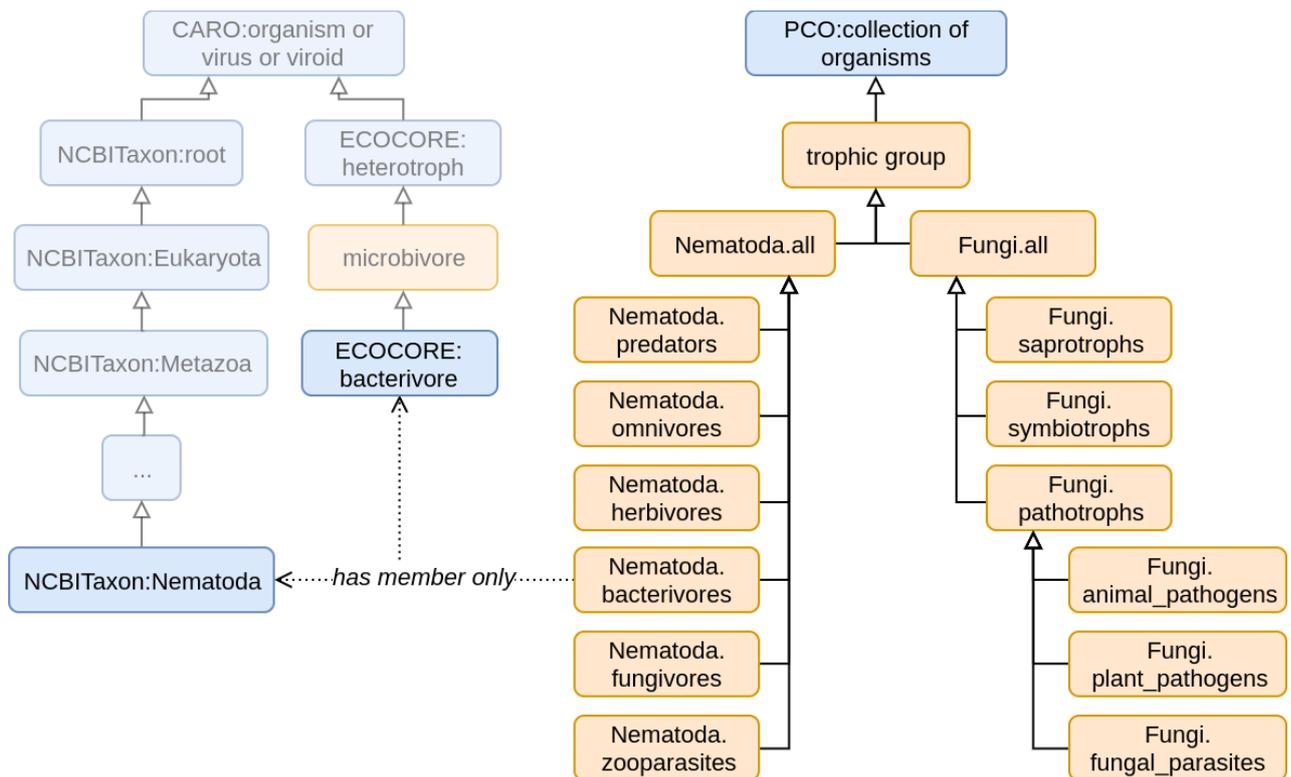
233

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

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

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

251



252

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

257

258

259 SFWO also provides logical definitions for most of these trophic group classes. These definitions

260 (“axioms”) introduce constraints on the type of organisms that can be members of each trophic group.

261 Examples of logical definitions for some nematode trophic groups are given in Table 1. The logical

262 definition of the class *Nematoda.bacterivores* [SFWO:0000275] states that members of this trophic

263 group are necessarily instances of the *Nematoda* and *bacterivore* [ECOCORE:00000098] classes.

264 These logical axioms allow a reasoner to automatically deduce the taxonomic class and diet of

265 organisms from knowledge of the trophic groups of which they are members (see subsection on

266 reasoning in SFWO).

267

268

269

270

271

Label	Logical definition
Nematoda.all	<i>trophic group</i> [SFWO:0000127] <i>and has member</i> [RO:0002351] <i>only</i> <i>Nematoda</i> [NCBITaxon:6231]
Nematoda.bacterivores	Nematoda.all [SFWO:0000273] <i>and has member</i> [RO:0002351] <i>only</i> (<i>Nematoda</i> [NCBITaxon:6231] <i>and bacterivore</i> [ECOCORE:0000098])
Nematoda.fungivores	Nematoda.all [SFWO:0000273] <i>and has member</i> [RO:0002351] <i>only</i> (<i>Nematoda</i> [NCBITaxon:6231] <i>and fungivore</i> [ECOCORE:0000015])
Nematoda.zooparasites	Nematoda.all [SFWO:0000273] <i>and has member</i> [RO:0002351] <i>only</i> (<i>Nematoda</i> [NCBITaxon:6231] <i>and zooparasite</i> [SFWO:0000279])

272

273

274

275

276

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.

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

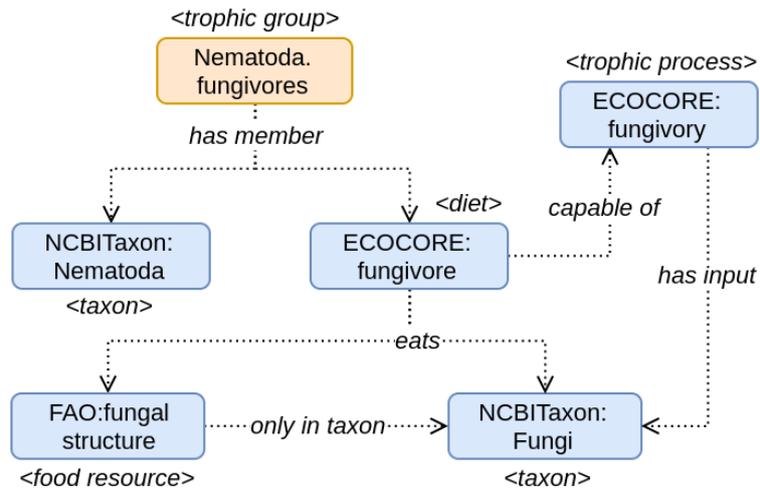
295

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.

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

306

307 *Trophic interactions*. Trophic interactions are relational concepts linking a consumer to a food source.
308 They are naturally represented in SFWO using object properties. Following best practices of ontology
309 development, SFWO reuses object properties from the Relations Ontology (RO). In particular, SFWO
310 imports an ecological subset of RO whose top-level object property *ecologically related to*
311 [[RO:0002321](#)] is the root of a branch that comprises many terms for representing biotic and trophic
312 interactions between individuals. The two terms most often used as part of the diet class logical
313 definitions are *eats* [[RO:0002470](#)] — e.g. in the definition of *nematophage* — and *parasite of*
314 [[RO:0002444](#)] — e.g. in the definition of *zooparasite*. Note that *parasite of* is not considered to be a
315 trophic interaction but a symbiotic interaction in RO, this latter being defined as “a biotic interaction in
316 which the two organisms live together in more or less intimate association”. In its current version, RO
317 is missing terms to represent symbiotrophic interactions in which one organism acquires nutrients
318 through a symbiotic relationship with another organism, e.g. trophic parasitism, mycorrhizal
319 associations, etc. The object property *symbiotrophically interacts with* [[SFWO:0000099](#)] was created
320 in SFWO as a temporary solution, and a new term request has been submitted to the GitHub issue
321 tracker of RO⁵.



322

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

325 Reasoning in SFWO

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

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

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

337 $\text{subClassOf}(\text{Nematoda} \text{ and } \text{fungivore}, \text{member of value } \text{Nematoda.fungivores})$

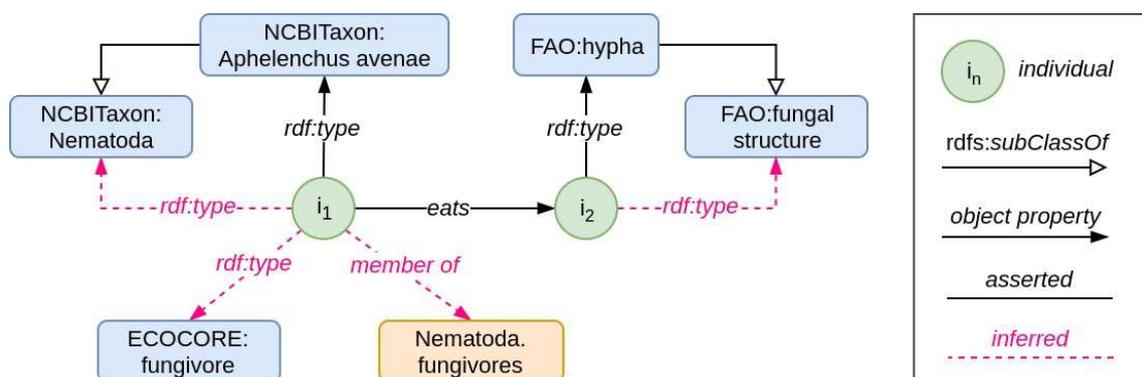
338 is a general class axiom stating that any individual organism that is an instance of the anonymous
 339 class (*Nematoda and fungivore*), i.e. the intersection of classes *Nematoda* and *fungivore*
 340 [[ECOCORE:00000015](#)], is also a member of the trophic group *Nematoda.fungivores*
 341 [[SFWO:0000276](#)].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

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

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

- 348 1. i_1 is a *Nematoda*, because *A. avenae* is a subclass of *Aphelenchus*, which is a subclass of
349 *Aphelenchidae*, which is a subclass of... , which is a subclass of *Nematoda* ;
- 350 2. i_2 is a *fungal structure*, because *hypha* is a subclass of *multicellular fungal structure*, which is
351 a subclass of *fungal structure* ;
- 352 3. i_1 is a *fungivore*, because i_1 eats i_2 and i_2 is a *fungal structure* and *fungivore* is logically
353 equivalent to *eats some (Fungi or fungal structure)* ;
- 354 4. i_1 is a member of the trophic group *Nematoda.fungivores*, from inferred facts 1 and 3 and the
355 general class axiom linking *Nematoda*, *fungivore*, and *Nematoda.fungivores*.

356
357



358

359 **Figure 7.** A reasoner is able to derive logical consequences from asserted facts using the ontology
360 axioms. Here, the diet and trophic group of *A. avenae* are inferred from the asserted fact that this
361 species of nematode feeds on hyphae.
362

363 Terminological issues

364 The development of SFWO has highlighted a number of terminological issues currently present in soil
365 ecology that were, at least partly, addressed here (see terminological discussions on SFWO's GitHub
366 issue tracker⁶). Below we describe a few of these issues as examples of the challenges in
367 harmonising the terminology in a scientific domain and modelling these concepts as part of a formal

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

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

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

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

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

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

426 SFWO's development and maintenance

427 The ontology design process started with a small team of experts in soil invertebrate ecology who
428 provided a list of over 100 terms related to primary food groups, diets and food resources of soil-
429 associated consumers, their definitions and relationships between diets and food resources, e.g.
430 *entomophage* → *Insecta* or *saprophylophage* → *dead wood*. A subset of these core terms was
431 manually aligned with classes from existing ontologies using Ontobee's search engine [20]. For
432 instance, the term "bacterivore" matched the class *bacterivore* [[ECOCORE:00000098](#)] defined in the
433 ECOCORE ontology. These classes were used as "seeds" for module extraction, a popular strategy
434 for ontology reuse that avoids the overheads involved in importing complete ontologies [21]. Modules
435 were merged to form the backbone of SFWO. New classes for the terms that could not be imported
436 from existing ontologies were created manually using Protégé⁷, a free and open-source ontology
437 editor. The ontology development workflow is managed and automated using the Ontology
438 Development Kit [22]. The ontology is under version control in a GitHub repository⁸ which is also used
439 to track changes and issues, and manage releases.

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

448 SFWO is continuously updated, improved and enriched thanks to contributors' suggestions for
449 revisions. Currently, SFWO brings together a community of a dozen soil ecology experts. Contributors
450 are invited to suggest and discuss changes in the ontology using the issue tracker of the project's
451 GitHub repository. Requests for changes/modifications are regularly submitted to the ontology
452 steering committee⁹ for discussion and validation. If approved, changes are applied using the above
453 methodology (extraction of modules or creation of new classes, manual editing to add synonyms or to

13 ⁷ <https://protege.stanford.edu/>

14 ⁸ http://github.com/nleguillarme/soil_food_web_ontology

15 ⁹ The composition of the steering committee can be found on the project's GitHub page.

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

458 Results

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

464 Standardisation of trophic trait datasets

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

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

481

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)

482 **Table 2.** Samples of the BETSI database of soil invertebrate traits [25]. These samples consist of
 483 dietary information for one genus and one species of Carabidae. Data in the attribute_trait column are
 484 matched with the corresponding concepts in SFWO.
 485
 486

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)

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

491 Trophic data integration and knowledge base construction

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

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

<i>subject</i>	<i>predicate</i>	<i>object</i>	
_: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	The identifier of the individual described in the first line of Table 2
_:betsi_carabidae_1	rdf:type	NCBITaxon:106379	
_:betsi_carabidae_1	rdf:type	ECOCORE:00000088	
_:betsi_carabidae_2	rdf:type	CARO:0001010	(1) This triple expresses the fact that _:betsi_carabidae_0 is an instance of <i>organism or virus or viroid</i> [CARO:0001010]
_:betsi_carabidae_2	rdf:type	NCBITaxon:106379	
_:betsi_carabidae_2	rdf:type	ECOCORE:00000090	
_:rainford_mayhew_0	rdf:type	CARO:0001010	(2) This triple expresses the fact that _:betsi_carabidae_0 is an instance of <i>Abacetus</i> [NCBITaxon:170239]
_:rainford_mayhew_0	rdf:type	NCBITaxon:60838	
_:rainford_mayhew_0	rdf:type	ECOCORE:00000089	
_:rainford_mayhew_1	rdf:type	CARO:0001010	(3) This triple expresses the fact that _:betsi_carabidae_0 is an instance of <i>carnivore</i> [ECOCORE:00000088]
_: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	

507

508 **Listing 1.** A semantic data integration approach such as the one described in [27] can transform data
 509 from multiple heterogeneous data sources into a common format — here, the Resource Description
 510 Framework (RDF) data model. The atomic unit of data in RDF is called a triple. A RDF triple
 511 expresses a relationship between two entities in the form of a subject-predicate-object expression. In
 512 this example, the annotated data in Table 2 and Table 3 have been transformed to RDF triples and
 513 combined into a single RDF dataset. This provides a unified view over the initial data sources that can
 514 be queried using the SPARQL Query Language¹⁰ for RDF to retrieve information across sources.

515 Soil food web reconstruction

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

522 broad spectrum of taxonomic groups, SFWO provides a common, logically-sound framework for soil-
523 associated consumers classification. This framework solves many issues related to inconsistent
524 definitions of trophic groups across published studies and the heterogeneity in food web aggregation
525 levels that they entail, thus making food webs reconstructed in different research contexts more
526 reusable, comparable and combinable as part of e.g., meta-analyses. In addition, SFWO's logical
527 axioms make it possible to automate the process of assigning taxa to trophic groups using a reasoner,
528 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 rdfs:type ?taxon.
  ?taxon rdfs:subClassOf NCBITaxon:1. } (2)
  ?taxon rdfs:label ?taxon_name. (3)
}
```

(1) Assign all individuals that are *member of* [RO:0002350] the trophic group *Coleoptera.detrivores* [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

529

taxon	taxon_name
NCBITaxon:106379	Abax ovalis
NCBITaxon:50548	Cupedidae

530

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

539

540

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

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

551 Trophic information extraction

552 The above applications of SFWO rely on our ability to compile large databases on the trophic ecology
553 of soil consumers. Although much data is available from previously published research, the creation of
554 literature-based datasets requires significant manual investment for literature searching, acquisition,
555 screening, data extraction, and harmonisation of entities, such as species or trait names [29]. An
556 alternative is to use information extraction approaches to automatically turn unstructured text into
557 structured data, an approach currently taken by the Specialised Information Service Biodiversity
558 Research¹¹ (BIOfid), which aims at extracting structured data from legacy Central European literature
559 through semantic role labelling. With its large and ever-increasing terminology related to soil trophic
560 ecology, SFWO is a valuable lexical resource for trophic information extraction. The most
561 straightforward way to use SFWO as part of an information extraction pipeline is to assemble all or
562 some of the terms defined in the ontology into a gazetteer, i.e. a list containing names of entities of
563 interest, and use this gazetteer to find mentions of these entities in text (Figure 8). SFWO is
564 particularly suitable to detect mentions of trophic processes, food resources and diets in text. In
565 combination with a taxonomic entity recognition system such as the one described in Le Guillarme
566 and Thuiller [30], SFWO could support the automatic creation, curation and updating of literature-
567 based trophic trait datasets.

568

In both forests soil animals could be classified either as **saprophages DIET**, including microphytophages, or **predators DIET**. On average, the $\delta^{15}\text{N}$ of **predatory DIET** taxa (**Chilopoda LIVB**, **Araneida LIVB**, **Gamasina LIVB**, **Staphylinidae LIVB**) exceeded that of **saprophagous DIET** or **microphytophagous DIET** taxa (**Lumbricidae LIVB**, **Isopoda LIVB**, **Diplopoda LIVB**, **Collembola LIVB**, **Oribatida LIVB**, **Enchytraeidae LIVB**) by 4.4 and 3.9‰ for the Göttinger Wald and the Solling, respectively. We assume that most of the **saprophagous DIET** or **microphytophagous DIET** taxa studied consist of primary and secondary **decomposers DIET** and hypothesize that **predators DIET** prey more on secondary than primary **decomposers DIET**.

569

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

574 Discussion and conclusions

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

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

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

598 Perhaps the most difficult aspect of developing an ontology is getting it accepted and adopted by the
599 scientific community. A few factors can facilitate engagement of the community of soil scientists. First,
600 the ontology must be consensual, which means that it must capture the knowledge of the domain in a
601 way that is accepted by the majority, if not by all. To help build and maintain consensus over the
602 years, we have designed SFWO's development as a collaborative process in which everybody can
603 participate, along the lines of open source software development. Therefore, on behalf of the steering
604 committee, we invite soil scientists, data providers, and anyone interested in using SFWO to
605 contribute to the improvement of the ontology through the project's GitHub repository. Furthermore, to
606 facilitate the adoption of SFWO and the dialogue between ontology developers and domain experts
607 (likely unfamiliar with the arcana of ontological languages such as OWL), we have made the ontology
608 available in more user-friendly formats. A browsable version of SFWO is available online on
609 AgroPortal¹², with the IRIs of the SFWO concepts all being redirected to AgroPortal. It also includes a
610 graphical representation of the concepts hierarchy, accessible from the "Visualization" tab in the
611 Classes view. In addition, each release of SFWO will be accompanied by an export of the ontology in
612 a tabular format (CSV file). The export file of the current version of SFWO accompanying this paper is
613 provided as Supplementary Material. Finally, we will provide comprehensive documentation so that
614 users can better understand the architecture of the ontology and the design choices that led to that
615 architecture.

616

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

622 data on soil trophic ecology. In addition, SFWO opens avenues for developing practical ontology-
623 based applications in the field. We also hope this work will foster other ontology development
624 initiatives addressing other aspects of soil ecology.
625

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 *zooparasite* 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.

626

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

627 Acknowledgements

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

637 References

- 638 [1] Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem
639 functioning. *Nature*, 515(7528), 505-511.
640
- 641 [2] Potapov, A. M., Beaulieu, F., Birkhofer, K., Bluhm, S. L., Degtyarev, M. I., Devetter, M., ... &
642 Scheu, S. (2022). Feeding habits and multifunctional classification of soil-associated consumers from
643 protists to vertebrates. *Biological Reviews*.
644
- 645 [3] Hedde, M., Blight, O., Briones, M. J., Bonfanti, J., Brauman, A., Brondani, M., ... & Capowiez, Y.
646 (2022). A common framework for developing robust soil fauna classifications. *Geoderma*, 426,
647 116073.
648
- 649 [4] Potapov, A. M. (2022). Multifunctionality of belowground food webs: resource, size and spatial
650 energy channels. *Biological Reviews*.
651

- 652 [5] Gauzens, B., Legendre, S., Lazzaro, X., & Lacroix, G. (2013). Food-web aggregation,
653 methodological and functional issues. *Oikos*, *122*(11), 1606-1615.
654
- 655 [6] Pey, B., Laporte, M. A., Nahmani, J., Auclerc, A., Capowiez, Y., Caro, G., ... & Hedde, M. (2014). A
656 thesaurus for soil invertebrate trait-based approaches. *PLoS One*, *9*(10), e108985.
657
- 658 [7] Hunt, H. W., Coleman, D. C., Ingham, E. R., Ingham, R. E., Elliott, E. T., Moore, J. C., ... & Morley,
659 C. R. (1987). The detrital food web in a shortgrass prairie. *Biology and Fertility of Soils*, *3*, 57-68.
660
- 661 [8] Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology
662 for describing and synthesizing ecological observation data. *Ecological informatics*, *2*(3), 279-296.
663
- 664 [9] Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., ... & Wooley, J. (2014).
665 Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections
666 ontology and related ontologies. *PloS one*, *9*(3), e89606.
667
- 668 [10] Hoehndorf, R., Schofield, P. N., & Gkoutos, G. V. (2015). The role of ontologies in biological and
669 biomedical research: a functional perspective. *Briefings in bioinformatics*, *16*(6), 1069-1080.
670
- 671 [11] Otte, J. N., Beverley, J., & Ruttenberg, A. (2022). BFO: Basic formal ontology. *Applied ontology*,
672 (Preprint), 1-27.
673
- 674 [12] Schoch, C. (2011). NCBI Taxonomy. 2011 Apr 7 [Updated 2020 Feb 11]. In: Taxonomy Help
675 [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2011-. Available from:
676 <https://www.ncbi.nlm.nih.gov/books/NBK53758/>
677
- 678 [13] Pölme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B. D., Clemmensen, K. E., Kauserud,
679 H., ... & Tedersoo, L. (2020). FungalTraits: a user-friendly traits database of fungi and fungus-like
680 stramenopiles. *Fungal Diversity*, *105*(1), 1-16.
681
- 682 [14] Shen, H., Shiratori, Y., Ohta, S., Masuda, Y., Isobe, K., & Senoo, K. (2021). Mitigating N₂O
683 emissions from agricultural soils with fungivorous mites. *The ISME Journal*, *15*(8), 2427-2439.
684
- 685 [15] Potapov, A. M., & Tiunov, A. V. (2016). Stable isotope composition of mycophagous collembolans
686 versus mycotrophic plants: do soil invertebrates feed on mycorrhizal fungi?. *Soil Biology and*
687 *Biochemistry*, *93*, 115-118.
688
- 689 [16] Jassey, V. E., Walcker, R., Kardol, P., Geisen, S., Heger, T., Lamentowicz, M., ... & Lara, E.
690 (2022). Contribution of soil algae to the global carbon cycle. *New Phytologist*, *234*(1), 64-76.
691
- 692 [17] Bolton, J. J. (2016). What is aquatic botany?—And why algae are plants: The importance of non-
693 taxonomic terms for groups of organisms. *Aquatic Botany*, *132*, 1-4.
694
- 695 [18] Moore, J. C., Berlow, E. L., Coleman, D. C., de Ruiter, P. C., Dong, Q., Hastings, A., ... & Wall, D.
696 H. (2004). Detritus, trophic dynamics and biodiversity. *Ecology letters*, *7*(7), 584-600.
697
- 698 [19] Swift, M. J., Heal, O. W., Anderson, J. M., & Anderson, J. M. (1979). *Decomposition in terrestrial*
699 *ecosystems* (Vol. 5). Univ of California Press.
700
- 701 [20] Xiang, Z., Mungall, C., Ruttenberg, A., & He, Y. (2011, July). Ontobee: A linked data server and
702 browser for ontology terms. In *ICBO*.
703

- 704 [21] Doran, P., Tamma, V., & Iannone, L. (2007, November). Ontology module extraction for ontology
705 reuse: an ontology engineering perspective. In *Proceedings of the sixteenth ACM conference on*
706 *Conference on information and knowledge management* (pp. 61-70).
707
- 708 [22] Matentzoglou, N., Goutte-Gattat, D., Tan, S. Z. K., Balhoff, J. P., Carbon, S., Caron, A. R., ... &
709 Osumi-Sutherland, D. (2022). Ontology Development Kit: a toolkit for building, maintaining, and
710 standardising biomedical ontologies. *arXiv e-prints*, arXiv-2207.
711
- 712 [23] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... & Lewis, S. (2007). The
713 OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature*
714 *biotechnology*, 25(11), 1251-1255.
715
- 716 [24] Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., ... &
717 Simons, N. K. (2019). Towards an ecological trait-data standard. *Methods in Ecology and Evolution*,
718 10(12), 2006-2019.
719
- 720 [25] Joimel, S., Nahmani, J., Hedde, M., Auclerc, A., Léa, B., Bonfanti, J., ... & Benjamin, P. (2021,
721 April). A large database on functional traits for soil ecologists: BETSI. In *Global Symposium on Soil*
722 *Biodiversity* (pp. 523-528).
723
- 724 [26] Rainford, J. L., & Mayhew, P. J. (2015). Diet evolution and clade richness in Hexapoda: a
725 phylogenetic study of higher taxa. *The American Naturalist*, 186(6), 777-791.
726
- 727 [27] Le Guillaume, N., & Thuiller, W. (2023). A Practical Approach to Constructing a Knowledge Graph
728 for Soil Ecological Research. *bioRxiv* 2023.03.02.530763; doi:
729 <https://doi.org/10.1101/2023.03.02.530763>
730
- 731 [28] Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015).
732 Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology,*
733 *Evolution, and Systematics*, 46, 523-549.
734
- 735 [29] Farrell, M. J., Brierley, L., Willoughby, A., Yates, A., & Mideo, N. (2022). Past and future uses of
736 text mining in ecology and evolution. *Proceedings of the Royal Society B*, 289(1975), 20212721.
737
- 738 [30] Le Guillaume, N., & Thuiller, W. (2022). TaxoNERD: deep neural models for the recognition of
739 taxonomic entities in the ecological and evolutionary literature. *Methods in Ecology and Evolution*,
740 13(3), 625-641.
741
- 742 [31] Eisenhauer, N., Bender, S. F., Calderón-Sanou, I., de Vries, F. T., Lembrechts, J. J., Thuiller,
743 W., ... & Potapov, A. (2022). Frontiers in soil ecology—Insights from the World Biodiversity Forum
744 2022. *Journal of Sustainable Agriculture and Environment*, 1(4), 245-261.