

Factors affecting ontology development in ecology

C. Maria Keet

KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano
Piazza Domenicani 3, 39100 Bozen-Bolzano, Italy, tel: +39 04710 16128

keet@inf.unibz.it

Abstract. Few ontologies in the ecological domain exist, but their development can take advantage of gained experience in other domains and from existing modeling practices in ecology. Taxonomies do not suffice because more expressive modeling techniques are already available in ecology, and the perspective of flow with its centrality of events and processes cannot be represented adequately in a taxonomy. Therefore, formal ontologies are required for sufficient expressivity and to be of benefit to ecologists, which also enables future reuse. We have created a formal mapping between the software-supported ecological modeling method and software tool STELLA and ontology elements, which simplifies bottom-up ontology development considerably and has excellent potential for semi-automated ontology development. However, the conducted experiments also revealed that ontology development for ecology is close to being part of ecological research that through the formalized representation of the knowledge more clearly points to lacunas and suggestions for further research in ecology.

1. Introduction

It is well-known that ontologies can be a valuable artifact for data(base) integration. However, for ontologies to be useful, one first needs to develop a good ontology that covers the domain accurately and precisely and has the right balance between utility and ontological correctness (the ontological trade-off). Although multiple engineering artifacts exist, from structured controlled vocabularies to formalized foundational ontologies, ontologies in the domain of ecology do not exist to the extent as, for instance, in cell biology. We can take advantage of lessons learned from developing ontologies in other biology disciplines, most notably in molecular biology and anatomy, and from suggestions made by philosophical ontologists. The former includes experiences with GO¹, OBO², and FMA³, the latter comprises the use of foundational ontological aspects like the nature of entities/concepts and (primitive) relations [1] and OntoClean [2] which provides a methodology for removing incorrect ontological decisions made in a taxonomy by relying on types of properties (characterising, sortal, phased sortal. etc.) and metaproperties (rigidity, identity, etc.). However, whether we can use a similar approach as taken by the Gene Ontology Consortium depends on the result of a comparative analysis between molecular biology and ecology (§2). One of

¹ Gene Ontology: <http://www.geneontology.org>.

² Open Biological Ontologies: <http://obo.sourceforge.net>.

³ Foundational Model of Anatomy: <http://sig.biostr.washington.edu/projects/fm/index.html>.

the differences is that there is an established practice of modeling in ecology, that, albeit different from computer science and ontology research, can be advantageous to enhance ontology development. A widely used, software-supported, ecological modeling technique is STELLA⁴, which we have exploited in formulating formal correspondences between STELLA model elements and ontology elements (§4). This was identified and put to the test with formalizing ecological knowledge contained in a STELLA pollution example and the Microbial Loop (ML) model [3], reported in §5 that also contains several ontology development considerations. Apart from simplifying and speeding up ontology development by using the formalization, related facets benefiting ontology development for ecology are discussed in §6 and the potential for semi-automatic bottom-up ontology development based on STELLA models is assessed. We finalize with some conclusions in §7.

2. Some Salient Features of the Ecology Subject Domain

An important factor in ecological and biogeochemical models is the *flow* of components in a eco(sub)system⁵, i.e. the path components take or sequence of processes it is involved in. A component can be a specific nutrient, such as nitrogen- or carbon-containing substances, pollutant, energy, and so forth, hence the centrality of endurants (entities that are wholly present in time) and instances thereof. However, the ‘component of concern’ is firmly embedded in the flow. For example, the nitrogen cycle from nitrate in soil to bacterium (nitrogen fixation by e.g. a *Rhizobium* sp.), transfer to a leguminous plant (like clover) with which the bacteria live in symbiosis with, transport within the plant, consumption by a ruminant, metabolism of the animal, excretion by animal, return of (some of) the nitrogen-containing molecules back to soil. One also can consider such cycles as a process of nested processes, i.e. from a perspective of a specific combination or sequence of distinct perdurants (entities that are partially present and happen in time). Thus the relation between ‘stuff’ (a substance, amount of matter etc.) and what happens to it are inextricably linked to one another. Conversely, molecular biologists do distinguish more clearly a separation between structural components, their functions and the processes in which they can be involved. GO consists of three distinct ontologies: Molecular Function (MF, describing activities), Biological Process (BP, with biological goals), and Cellular Component (CC, for locations) [4]. This approach treats perdurants *as if* they are endurants, but this objectification does not solve the connection between endurants and perdurants. For example, if one wants to couple some biological process with a cellular component, new relationships between the two ontologies need to be created (e.g. [5]). Thus, adding new knowledge about the combination that may result in a separate new ‘situation ontology’, or a mapping ontology that is positioned between BP and CC. However, that ecological modelers use tightly coupled endurants and perdurants does not necessarily prevent an ontologist to create artificial divisions between the two.

⁴ ISEE Systems: <http://www.iseesystems.com>; *ithink* is the same tool but used for business modeling.

⁵ For the remainder of the article, ‘ecological model’ comprises both types – a biogeochemical model is element-conserving, but this aspect is irrelevant for ontology development.

Perdurants include types of entities such as processes, events and states, in contrast with modeling paradigms in informatics and most ontological investigations, where the center of attention is the entity of the thing-quality paradigm. Philosophically, there are arguments for and against such emphasis: processes can only exist when there are endurants that are the ‘carriers’ of the process [5]. On the other hand, objects only come into existence through a process (refer to [6] and [7] for a wider scope of arguments). Few agreed-upon ontological categorizations exist, as can be observed in *Figure 1* or the Process Specification Language⁶, Business Process Management Initiative⁷, and Petri-nets. From an ontology engineering perspective, the approaches vary. One tactic is to separate perdurants from endurants linked by a participation relation as in DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [8], [1]. The Basic Formal Ontology (BFO⁸) consists of SNAP and SPAN ontologies where the latter includes a time perspective. However, none addresses the thing-process aspect as fully *interdependent*, which poses a potential problem when representing ecological knowledge in an ontology. Bittner *et al* [9] go to some length in formalizing the difference between enduring and perdurant, but this does not solve the nature of the relation when viewed from different perspectives. The Standard Upper Ontology has set up a 4D Ontology Working group⁹, without useable results as of yet.

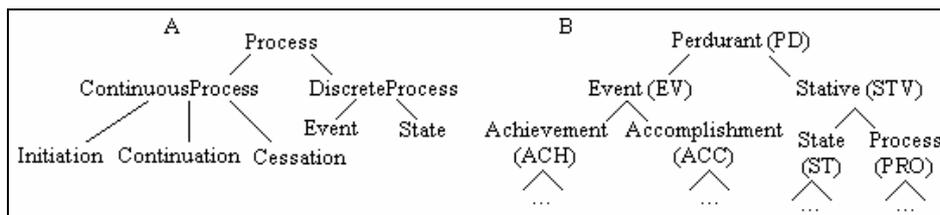


Fig. 1. Two examples of process-related categorizations. A: [6]; B: DOLCE [1].

A second difference lies in the level of granularity and demarcation of the discipline. The boundary of molecular biology lies at the cell-level and smaller entities, whereas in ecology ‘ecosystem’ and ‘the environment’ have fiat boundaries. On the one hand, earth is one ecosystem, but also the Amazon basin or the micro-environment in tree tops. Other methods of differentiation include trophic levels and ‘grand processes’ such as eutrophication and El Niño. While there are different ways of partitioning the domain at the molecular biology level, such scope in ecology is wider, thus when developing an ontology it requires involvement of a broader range of (sub-)disciplines that are less reductionist than molecular biology.

Third, ecologists have a different starting position compared to molecular biologists when modeling domain knowledge. Whereas in molecular biology no established modeling tradition existed, ecologists do have multiple established standards such as Odum’s conventions [10] and STELLA that are, depending on the sub-discipline in ecology, more or less often used. STELLA is relatively widely used and has software support comprising graphical elements and automatic generation of equations. STELLA is used in research and education for systems analysis and simu-

⁶ <http://www.mel.nist.gov/psl/index.html>.

⁷ <http://www.bpmi.org/>.

⁸ <http://ontology.buffalo.edu/bfo/>, http://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf.

⁹ <http://suo.ieee.org/SUO/SUO-4D/index.html>.

lations of, for example, predator-prey interaction, effects of contamination, and food chains. Ontologists should take advantage of these models for bottom-up ontology development. However, this also means that one cannot begin with a structured controlled vocabulary: whatever ontology is developed has to surpass semantic expressivity of existing ecological models for it to be perceived to be of benefit to ecologists.

A preliminary experiment in ontology development for ecology was carried out with a simplified pollution example, which confirmed that an informal ontology limited to *isA* and *partOf* relationships could not capture the expressivity of its corresponding STELLA model. The “extended semantic representation of equations” via “placeholder objects” [11] did not represent the (partially implicit) semantics of the STELLA model fully either (results omitted). An additional advantage of using STELLA as a starting point for ontology development is that, with the mapping between STELLA and ontology elements, the STELLA representation serves as an intermediate representation. Thereby it bridges the two disciplines with a common ground for communication. This will speed up bottom-up ontology development, which will facilitate data integration sooner rather than later.

3. Methodology

The first experiment was carried out with STELLA v8 for Windows from ISEE Corporation and the demonstration model *Amalgamated Industries*. The abstraction of this model, including the STELLA terminology and modeling elements, was matched to ontology jargon. Protégé v2.1.1 with OWL Plugin v1.2 Beta (Build 139) was used to improve the level of formalization and test the translation. Racer v1.7.21 was used for the ‘classify taxonomy’ and ‘check consistency’ features; Graphviz v1.12 plug-in to activate OWLViz, and ezOWL plug-in (v20040412) were added for ontology visualization of the developed ontology. To test the translation between STELLA and ontology elements, we used the larger ML model (*Figure 4*), converted it into a list of candidate entities and relations, which was structured into a formal ontology, also in Protégé. Development of both the *Pollution* and *MicrobialLoop* ontologies was aided by structuring the candidate entities and relationships adhering to the formalized DOLCE foundational ontology, which is intended for making already formed conceptualizations explicit (refer to [1] for explanation and categories).

4. Abstractions and Matching

Before addressing the formalization, a small STELLA model (*Figure 2*) of the simplified pollution scenario is outlined for illustration. This model captures a scenario where a factory disposes toxic waste in the river that flows into the pond downstream, in turn killing organisms living in the pond depending on the pollutant concentration. The ecological ‘concept of concern’ is the concentration of the pollutant in the pond, which has the related influencing factors modeled ‘around’ it, such as the released amount of pollutant by the chemical plant. There are three main aspects: water and pollutant in/outflow of the bound system, the combination of water volume and amount of pollutant determining the pollutant concentration in the pond, and the com-

bination of water outflow and pollutant concentration determines the amount of pollutant outflow. There are two factors of interest in comparing this type of model with its variants in computing, such as UML class diagrams, (E)ER and ontologies:

1. The ecological model is *event centered*, hence contains the representation of time, diagrammatically represented with the horizontal thick arrows with an open shaft, or phrased as the *route* taken by an element.
2. Key aspects in the ecological model are Flow, Stock, Converter, and Action Connector. A Stock correspond to a noun, being it particulars or universals, Flow to verb, Converter to attribute or property related to Flow or Stock, and Action Connector relates the former. *Figure 3* contains the comparison with computing verbiage (top half). *Object* is a candidate for an entity, *event_or_activity* in OO terms a candidate for a method and in an ontology categorised under a perdurant hierarchy and converter maps to *attribute_or_property*, which says something about the object, such as the outflow rate. The Action Connector (thin line with arrow) may be candidate for binary (ternary?) relationship between any two of Flow, Stock and Converter.

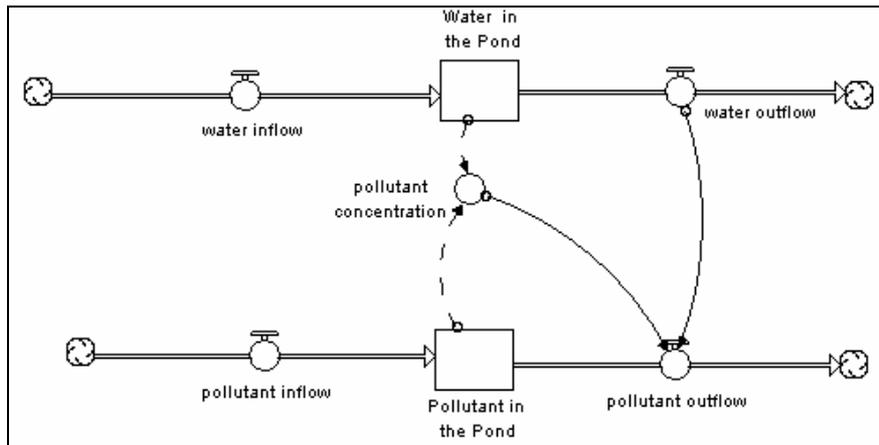


Fig. 2. Abstraction of the pollution example.

Following from 2), the formalization for the translation is:

$$\forall x ((\text{Stock}(x) \leftrightarrow \text{Entity}(x)) \rightarrow \text{ED}(x)) \quad (1)$$

$$\forall x ((\text{Flow}(x) \leftrightarrow \text{Entity}(x)) \rightarrow \text{PD}(x)) \quad (2)$$

$$\forall x ((\text{Converter}(x) \leftrightarrow \text{Entity}(x)) \rightarrow (\text{Q}(x) \vee \text{ST}(x))) \quad (3)$$

$$\forall x (\text{ActionConnector}(x) \leftrightarrow \text{Relationship}(x)) \quad (4)$$

where ED means endurant, PD perdurant, Q quality ('attribute' belonging to an entity), and ST state. Axiom (3) is open to experimentation: for example, the Converter *Pollutant concentration* in *Figure 2* can be a quality of the pond, liquid mixture, or detritus – anything that can be polluted – but also a state as in 'the pollutant concentration of the pond where the river enters' or 'the pollutant concentration of the pond

on 20-7-2004'. Further, *Pollutant concentration* may be subsumed by *Concentration* that can be applicable to a wider range of endurants and as such is not necessarily an *essential* property (a pond is still a pond with or without some concentration of a pollutant) but a situational one, or having concentration as a non-rigid property. Therefore, (pollutant) *Concentration* is better modeled as an ST, but at this stage of the investigation, the mapping of Converter to Q cannot be excluded with certainty. This ambiguity will be resolved by applying the proposed formalization to a formal pollution ontology and the larger ML model, which will clarify if the mappings are correct, shed light on the distribution of Q and ST from a Converter, and might be solved by adding additional axioms taking into account the context of the STELLA elements, such as how the converters are related to the other elements.

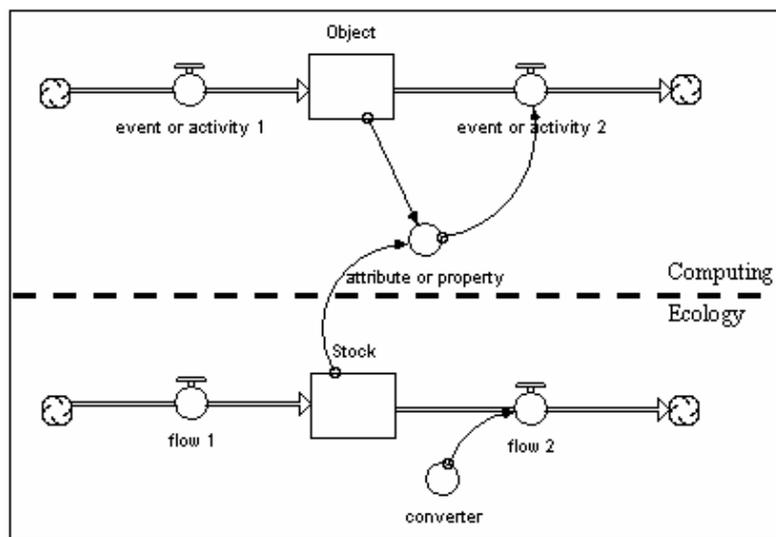


Fig. 3. Comparing the ecological model components with its analogue in a computing model.

The consequences of translating an ecological model into an ontology based on the provided mapping is that temporality and the movement of energy or nutrients is not yet represented exactly as is captured in the ecological model apart from incorporating the fact that types of processes exist; however these extras in STELLA are epistemological aspects. The original ecological model now can be remodeled into an ontology consisting of three features: endurants, 'attributes', and perdurants; what remains to be solved are the relations between them, i.e. the Action Connectors. With further testing of larger STELLA models such as LEEDS (Lake Eutrophication, Effect, Dose, Sensitivity model) and ML, and the provided formalization, it is possible to generate an ontology by 'loading' several of the STELLA ecological models into one of the ontology development tools.

5. Ontology Development

5.1 A Pollution Ontology

5.1.1 Motivating Example: Some Issues with an Informal Ontology

An informal ontology restricted to concepts and the isA or partOf relation does not suffice. For example, an isA relationship between *Water* and *Molecule*: although *Water* is indeed a molecule, *Water in the context of some ecological site* is not pure H₂O, but water containing dissolved molecules and suspended particles, i.e. water as a mixture (an amount of matter). The methodology of ontology base & commitment layers of DOGMA [12] may be more advantageous, because *Water isA Molecule* can be included in the ontology base and omitted from a commitment layer for an ecological site, whereas it would be included in a commitment layer of a chemicals ontology that omits *Water isA LiquidMixture*. In a simple taxonomy these options are unavailable.

Volume, *Rate* and *Concentration* capture a characteristic of their respective entity they are attached to, alike an attribute. *Molecule hasA Concentration*, but it can only have a concentration dissolved or suspended *in something* and not *of itself*; of itself are properties like melting temperature and structure of a molecule. However, to conclude it is an attribute or state of only water in the pond is premature: if modeled as such, the model will be unable to accommodate pollutants in sea, air, soil etc.

5.1.2 Upgrading to a Pollution Ontology

It is sub-optimal if one has to use different types of representation models (multiple taxonomies and placeholder objects) to capture the semantics. For a formalization to be exploited in full, one needs a formal ontology and a tool that is based on logic. Two widely used ontology development and editing tools are DAG-Edit and Protégé; the former provides functionality for structured controlled vocabularies (and taxonomies), whereas the latter is frame-based with Description Logic (DL) and OWL support. The DL version supports “maximum expressiveness without losing computational completeness ... and decidability ... OWL DL includes all OWL language constructs with restrictions such as type separation (a class can not also be an individual or property, a property can not also be an individual or class)”¹⁰. DAG-Edit is insufficient for the task, because relationship types are limited to isA and underspecified partOf (sometimes also developedFrom). Protégé, on the other hand, supports reasoning and allows higher expressivity by allowing specification of e.g. properties, range restrictions, and disjointness of entities. However, it also supports multiple inheritance, which complicates inferencing over the ontology and multiple inheritance may indicate bad modeling decisions, but this is not necessarily so.

Multiple inheritance. In the initial categorization, *Molecule* directly subsumed *Pollutant*, *Nutrient*, *OrganicMolecule*, and *AnorganicMolecule*, where some molecules, like *PHB* and *Phosphate*, were subsumed by *OrganicMolecule* and *Pollutant*, and *Nutrient* respectively. Biologically, these are correct statements, but it would be better to specify (not possible in Protégé, but implementable in a DL knowledge base):

if (concentration of AnorganicMolecule x) in (WaterBasin or AmountOfMatter) is higher than [some number] then (x isA Pollutant and x isA AnorganicMolecule) else (x isA Nutrient and x isA AnorganicMolecule) for species y.

¹⁰ <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.

This statement indicates that the difference between pollutant and nutrient is problematic: pollutants are harmful and nutrients beneficial to organisms¹¹, thus a functional categorization. However, there are two obstacles: first, a pollutant for species x can be a nutrient to species y ; this information may be used for *in situ* bioremediation¹², hence lies within the UoD. Secondly, low molecule concentration can be a nutrient but excess concentration pollutant; but when is ‘excess’ concentration high enough to classify it as pollutant? Narrowing down *Nutrient* to *NutrientBioremediation* prevents confusion with generic nutrients that never function as nutrients for bioremediation. That *Molecule* subsumes *Pollutant* and *Nutrient* together with *OrganicMolecule* and *AnorganicMolecule* is incorrect, because the former are *functions* assigned to the molecules, whereas the distinction for *OrganicMolecule* and *AnorganicMolecule* is based on the *composition* of molecules. Using the DOLCE categories for guidance, *Pollutant* and *Nutrient* still are physical objects (POB), but classified according to other criteria. For brevity of this experiment, they are separated into structure and the function of molecules. H_2O and *Amylose* each had only one isA relation hence were removed, but *PHB* and *3-chlorobenzoate* can be used for bioremediation as each one has 3 isA relationships: being an *OrganicMolecule*, *Pollutant*, and *Nutrient*. Although multiple inheritance has not been eliminated, due to having structure and function in one ontology, there is a major advantage in maintaining this sort of multiple inheritance: when one adds a new entity under *MoleculeStructure*, *Pollutant*, and *Nutrient*, i.e. the new entity has three isA relationships, then one can deduce there is potential for *in situ* bioremediation (as is the case with *3-chlorobenzoate*). On the other hand, if the new entity has two isA relationships, one to *MoleculeStructure* and the other to *Pollutant*, an ecosystem disruptive method to clean up the contaminated site is required; if there is only one isA relationship, then there is no need for the molecule to be in the classification because it does not serve any particular purpose for the UoD, or still needs to be grouped under *NutrientBioremediation* or *Pollutant*, thereby missing essential knowledge in the ontology. Thus from that perspective, multiple inheritance is not a ‘bad thing’ and can be used to derive additional information from querying the ontology.

Other modeling considerations and limitations include 1) Protégé prohibits creating an entity or individual starting with a number, in this case *3-chlorobenzoate*, which is problematic because names of many chemicals start with a number and appending the number at the end is not an option with more complex chemical structures. 2) The STELLA pollution model assumes *Pollutant_concentration_in* = *Pollutant_concentration_dumped_by_plant*, thereby ignoring adsorption and absorption to particulates in the river and sedimentation, consumption by organisms, and assuming that the pollutant concentration is uniform throughout the pond. Adsorption and absorption can be added to the ontology, but this extension is omitted from *Pollution* because its purpose is explorative with relation to the axioms. 3) Protégé tolerates unconstrained property creation, which can become prohibitive if one desires to develop an ontology with possible future ontology integration while another ontology has been

¹¹ Pollutant = “waste matter that contaminates the water, air or soil” (Wordnet) although nature may cause pollution as well; nutrient = “any substance that can be metabolized by an organism to give energy and build tissue” (Wordnet).

¹² *In situ* bioremediation: removing a pollutant from a contaminated site without disrupting the ecosystem by using organisms instead of soil excavation and chemical decontamination.

restricted to a few relationship types. 4) In order to create a sound ontological basis of the categorization of entities, the structure of the DOLCE top-level categorization was used, most notably the amount of matter (M), physical object (POB), and process (PRO) versus ST, resulting in 56 entities for the *Pollution* ontology. Whereas relating entities other than isA or partOf is not possible in a taxonomy, this is possible with a formal ontology and therefore included in *Pollution* via 9 properties and corresponding constraining axioms. The plug-in module OWLViz for Protégé only shows the isA relationships in the graphical representation; for additional expressivity, ezOWL is required. With a very small ontology, this creates a manageable view of the semantics, but even with only 56 entities, the diagram is already too large to be manageable (the OWL file of *Pollution* is available online as supplementary material). Comparing this ontology with the original STELLA pollution model, the 11 elements are ‘translated’ into 56 entities and 9 properties. This may seem excessive, but the ontology captures more semantics than its STELLA counterpart does, hence has a higher likelihood of being useful for more pollution models than STELLA’s dump-river-pond scenario. From the perspective of semi-automated bottom-up development of ontologies based on STELLA models, this poses a challenge: how labor-intensive is the additional structure one needs to add to adhere to sound ontological principles? Is it sensible to develop semi-automatic translation software if a considerable amount of ontology development effort may have to be carried out manually anyway? Instead of generating a structure of the ontology, a viable option is to translate STELLA elements into a list of entities and relationships that one needs to include in the ontology. This reduces the manual analysis because it is possible to develop a backbone domain ontology, ‘hang in’ the entities generated from the STELLA model, and augment this with the relationships and properties that resulted from the translation.

5.2 The Microbial Loop Model

The formalization was applied to the ML model (*Figure 4*) to examine if the axioms still hold in a real and larger STELLA model, to shed light on the distribution of Q and ST from a Converter, and to investigate if additional axioms are required when taking into account the context of the STELLA elements, such as how converters are related to other elements. ML’s initial mapping to ontological categories contain 38 STELLA elements, of which 11 Stock/ED, 21 Flow/PD, two Converters that map to ST, and four Action Connectors/Relationships that are modeled as properties in Protégé (mappings included as supplementary material). All Stock elements can be further categorized as Non-Agentive Physical Object (NAPO) leaf categories. Further, to accommodate these NAPOs in an ontology, extra entities related to the NAPOs were added, such as *Phytoplankton* (which is an Agentive Physical Object APO), and *Detritus* (an amount of matter M). Note that “Phyto C” is the organic carbon *component* of phytoplankton, not the phytoplankton as a whole. To accommodate for this in the ontology, adding phytoplankton only as an APO is insufficient. Apart from the phytoplankton carbon and nitrogen, the NPK parameters (Nitrogen, Phosphor, Potassium) are relevant for agriculture and soil science in particular. Should one include other molecules to be more comprehensive? From an ontological viewpoint probably yes, but one might argue a utilitarian restriction “it’ll do” for the intended purpose.

A separate issue concerning categorization of organisms in the ML as APOs emerged during a conversation with one of the creators of the ML model, Professor

Seventeen of the 21 Flow elements are processes (PRO) and the other three accomplishment (ACC) (2x the entity *Uptake* and *Excretion*) and achievement (ACH) (*Oxidation*). There is no particular contextual aspect why these three have to be categorized under a different type of perdurant. The two Converters, both called “grazing pressure”, each is a parameter of the process *Grazing*. Hence, it might be a quality of the process of grazing, because it is specifically constantly dependent on the entity it inheres in (grazing): at any time, a quality cannot be present unless the entity it inheres in, in this case a PD, is also present, and a PD is present if its ED bearer is present. However, if there is no plankton, the grazer (ED) may be grazing on something else. If there are no grazers, then grazing (PD) is not there and its grazing pressure as Q is also absent. Alternatively, the grazing pressure simply reaches zero, resulting in “grazing pressure” as a ST: the notion of “grazing pressure” is there, thus a ST and not a Q, which adheres to the ecology interpretation. Thus, this does not imply that Converters automatically always can be translated into states. The Action Connectors map well to properties (relationships between entities). There are 59 entities and 10 properties in the *MicrobialLoop* ontology (the OWL file is available as supplementary material), of which a summarized section is shown in *Figure 5*. For instance, the left-most rectangle is a visual representation of *Protozoa* as subtype of *Microorganisms*, with (5) inherited from *Microorganisms*, (6) and (7) as necessary properties of *Protozoa*, and (8) a necessary property of *PhytoPlankton*. *MicroAlgae* and *MacroAlgae* are disjoint subtypes of *Algae*, and *Phytoplankton* and *ZooPlankton* are disjoint subtypes of *Plankton*, and so forth.

$$\forall x \exists y (\text{Protozoa}(x) \rightarrow \text{hasProcess}(x, y) \wedge \text{Respiration}(y)) \quad (5)$$

$$\forall x \exists y (\text{Protozoa}(x) \rightarrow \text{grazesOn}(x, y) \wedge \text{PhytoPlankton}(y)) \quad (6)$$

$$\forall x \exists y (\text{Protozoa}(x) \rightarrow \text{grazesOn}(x, y) \wedge \text{Bacteria}(y)) \quad (7)$$

$$\forall x \exists y (\text{Phytoplankton}(x) \rightarrow \text{accomplishes}(x, y) \wedge \text{Uptake}(y)) \quad (8)$$

The translation axioms provide an accurate high-level mapping for both the *MicrobialLoop* and *Pollution*, although the amount of Converters and Action Connectors in the models may be to be too few to statistically conclusively confirm correctness of the mapping.

6. Discussion

Additional entities had to be added to the ontology compared to its STELLA variant. In spite of this, several factors ameliorate this issue. Ecology already divides concepts into three types: natural, functional and integrative concepts. The first two types of ecological concepts can be identified in the ontology: the functional concepts are categorized under *Perdurant* and the natural concepts subsumed by *PhysicalEndurant*. Imposing a separation and categorization may actually benefit ecology. Ford [13] presents the interdependencies between the three types of concepts indicating that “[n]ew functional concepts arise to describe newly understood structures or interactions in natural concepts and research into functional concepts is constantly used to refine the definition of existing natural concepts and their classifications” and “[d]evelopments in measurement lead to refinements of functional concepts”. Hence, by defining the concepts more clearly with the aid of formal ontology, the discipline

tology browsing from different perspectives exploiting a theory of granularity applied to the subject domain. Aside from diverging ideas on theories, the ‘windows on reality’ differ depending on what the scientist is looking for. In ecology, it is common to start with flows as opposed to starting with the object where processes act upon. From a formal ontological perspective, this is not necessarily problematic: creating the ontology starting with perdurants and subsequently axiomatizing their influence on endurants is possible. In addition, two distinct methodological approaches in ecological research exist. In theoretical ecology, one devises a theory that is tested on its applicability in nature afterwards. On the other hand, ecological modeling via the empirical approach involves tweaking the model until it fits the observed data, where only a limited set of parameters of the subject matter is used [16]. The former approach indicates that one starts with a framework that will be filled up inwardly and more densely, where the latter starts small and gradually will evolve by spreading outward once more research has been conducted. If one methodology is better than the other is irrelevant here, however, it is important to realize that both approaches are used in ecology, and, at least initially, might not result in the same output due to divergent design decisions. Also, cooperating with domain experts of either type involves a different approach commencing ontology development.

Although engineering challenges of ecology ontology development can be solved, the philosophical formal ontology development entertains itself with *what* and *how* to represent what is known, where an ontologist for ecology will have to cooperate in the process that otherwise logically occurs before ontology creation, i.e. the *why* in the semantics as part of regular science. The output must not only be usable for computer science (the ontology), but also of use to ecologists, who will be pushing the boundaries of their discipline by clarifying relevant concepts, thereby better formulating research questions, and later their theories. Provided alternative views of integrative concepts and theories can be accommodated for, it will aid the advance in ecological science. Apart from the difficulty on class/population (§5.2), a second aspect emerged during discussions with Tett: to compare and contrast more precisely a to-be-developed ontology of the STELLA model for the ‘model organism’ ERSEM¹³ with his *MicrobialLoop*. In addition, this ERSEM-ontology or *MicrobialLoop* might function as template or backbone for other ecological models in marine science. Another suggestion how ecology can benefit from the ontological approach is during the “fitting stage” of simulation models to match empirical data, where, according to Tett, parameters are added and removed “arbitrarily” and their values changed to fit observations. Adding the reasoning power of ontologies can ensure consistency. Inconsistencies introduced during the fitting stage provide a focus for (re)assessment and investigation of (a section of) the domain.

Considering some practical aspects of ontology development, the mapping between STELLA and ontology elements do not imply these correspondences will always be applicable, although the devised correspondences were confirmed to be sufficient for the *MicrobialLoop* development experiment. Accommodating the Converters remains less straightforward, because decisions have to be made to translate it to a quality or state. The use and meaning of the Action Connectors aided in determining the properties and relations in the *Pollution* and *MicrobialLoop*. The relative absence of serious difficulties during the modeling of characteristics of the biological entities may be due

¹³ European Regional Seas Ecosystem Model; refer to [17] for the structure and methodology of ERSEM and [18] for the microbial food web in marine systems.

to the size of the (randomly) chosen models and/or the author's domain knowledge. Initial challenges during the development of a taxonomy were absent during the 'upgrade' to *Pollution*, because the expressivity and flexibility of DL is much greater than the limited hierarchies in a taxonomy. Our experience confirms observations made by many other researchers that more expressive modeling languages *do* capture a richer semantics. This is not only because it compels the user to do so but also because one has the possibility to 'squeeze in' more knowledge, which in turn enforces closer inspection of the domain, resulting in ontologies with less errors and higher precision, hence are more stable. This is in contrast with e.g. DAG-Edit or standard UML class diagrams when one can gloss over such details. Moreover, where the flow dynamics cannot be addressed in a taxonomy, this is dealt with in the developed ontologies by first categorizing the relevant perdurants under *Process* and *State* and using properties to create the relationships between these entities and the endurants they act with/upon, all captured within *one* ontology instead of different representations. An alternative considered was BFO. However, developing two ontologies (SNAP and SPAN) that need to be 'connected' to capture the ecological semantics is prohibitive. The DOLCE top-level categories intuitively make sense and aids understanding of how distant or close biological semantics is from ontologies with a cognitive bias.

The two developed ontologies include more entities and relations than their STELLA counterpart and easily can be further extended to increase reusability. The latter can step up efforts to resolve ambiguities and assumptions; a very large ontology may be practically difficult to work with or requires full support of many sub-disciplines within ecology, analogous to the GO project [4]. Alternatively, one can take advantage of the extant modularization of ecological models: ML and e.g. Sea-Weed are composed of smaller sub-models, where the former contains Riley+, MicroPlankton and Autotroph-Heterotroph [3], and the latter Vollenweider¹⁴ and a tide & light simulation. Thus it is conceivable create 'mini-ontologies' based on the same foundational ontology principles for each ecological model separately, then develop a library where the user can choose the desired sections to create larger models, supported by a backbone ontology where for each situation one or a few ontologies can be attached to it. This approach also facilitates representation of diverging views of integrative ecological concepts. Challenges are the development of a backbone ontology and prospects of integrating ontologies. Analysis of differences between 'simple' and 'complex' versions of the Vollenweider models revealed that the more complex models contain both *additional* sections as well as filling the existing structure with more *detail*, thus both coverage and granularity changes. Tett and Wilson [3] indicate that this may be the case with multiple ecological models, because there is a desire to keep the amount of Stock elements to a minimum for reasons of computational power and practical as well as theoretical challenges of estimating parameters. Smith [19] claims that, in ecology terms, good *simulations* should include as much detail as possible, whereas good *models* should include as little as possible to capture the most important factors. This will affect ontology development for ecology, depending on with which kind of model/simulation one starts ontology development. Perceptions and knowledge about the subject domain change over time, is not always consensual, and have the potential destabilizing effect of cascading uncertainties to larger modular simulations and models, which are, according to [20], neither possible nor desirable to

¹⁴ <http://tejo.dcea.fct.unl.pt/resources.asp>. Vollenweider models form the basis for eutrophication control, which in turn is used in the LEEDS model and semantically related to ML.

include in one model. A design decision about one larger ontology versus multiple mini-ontologies will have to be made.

Concerning the *MicrobialLoop*, one may argue that the author's knowledge of the subject domain prevented the need for making excessive amounts of assumptions, such as knowing what "Phyto C" is, and microbiology in general. The outcome likely would have been different without such prior domain knowledge. Temporal factors such as accommodating changes in the rate of in/outflow are not addressed fully, because they are in the realm of instances. The richer expressiveness of the formal ontology approach using Protégé and DOLCE categories proved to be flexible enough for the task as it allowed correct representation of entities from taxonomies, entities that emerged from the semantic representation of equations, and other implicit knowledge of the STELLA models. The (untrained) ecologist indicated that the richer ezOWL graphical representation (*Figure 5*) that includes properties and constraints was preferred over a simplified taxonomic tree generated with OWLViz. Further, and more importantly, the ecologist judged the logic made the knowledge captured in the formal ontology become *clearer* than both STELLA and natural language, and considered to have useful potential to disambiguate the semantics to advance ecological research. The translation of the STELLA models into ontologies did introduce many new concepts, especially with the simple pollution experiment, but this was much less the case with the larger ML model. This indicates that with larger ecological models the issue of manual intervention during ontology development decreases. The translation axioms simplified ontology development from the ML considerably (a first version including initial mappings including comments was created within three hours and required only a few minor changes afterwards); therefore, utilizing other STELLA models with the provided formalization of the translation will also speed-up the overall development process of ontologies in ecology.

7. Conclusions

Although few ontologies in the ecological domain exist, their development can take advantage from existing modeling practices in ecology in particular. Taxonomies are insufficiently expressive compared to existing ecological modeling techniques and the perspective of flow in ecological models cannot be represented adequately in a taxonomy. We have created a formal mapping between the software-supported ecological modeling method STELLA and ontology elements, which simplifies bottom-up ontology development and has excellent potential for semi-automated ontology development. We developed two formal ontologies, *Pollution* and *MicrobialLoop*, in Protégé, exploiting the expressivity of OWL DL to capture the semantics of 'flow' in salient in ecology models. STELLA serves as an intermediate representation, widely used by ecologists and is translatable to a representation usable for ontologists. In addition, the more comprehensive semantics of the ontologies have not only a higher level of reusability within the domain, but also for future ontology integration as both *Pollution* and *MicrobialLoop* were developed with the same ontological foundational principles which shall facilitate reuse of ontologies. However, the experiments also revealed that ontology development for ecology is close to being part of ecological research that through the formalized representation of the knowledge more clearly points to lacunas and suggestions for further research in ecology, thereby aiding hy-

pothesis generation. We are currently extending this research with ontology development and management aspects such as modularization and ontology integration.

Acknowledgements

We wish to thank Paul Tett for his contribution to discuss the *MicrobialLoop* & ML. Some of the research was carried out while CMK was at Napier University, UK.

References

1. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: WonderWeb Deliverable D18 – Ontology library. WonderWeb. <http://wonderweb.semanticweb.org/> (2003)
2. Guarino, N., Welty, C.A.: An overview of OntoClean. In: Handbook on ontologies. Staab, S., Studer, R. (eds.). Springer Verlag, Berlin (2004) 151-159
3. Tett, P., Wilson, H.: From biogeochemical to ecological models of marine microplankton. *J. Mar. Sys.* 25 (2000) 431-446
4. Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource. *Nucl. Acids Res.* 32 (2004) D258-D261
5. Fielding, J.M., Simon, J., Ceusters, W., Smith, B.: Ontological theory for ontological engineering: biomedical systems information integration. 9th International Conference on the Principles of Knowledge Representation and Reasoning (KR2004), Whistler, Canada (2004)
6. Sowa, J.F.: Signs, Processes, and Language Games – Foundations for Ontology. <http://www.jfsowa.com/pubs/signproc.htm>. Date accessed: 16-6-2004. (2003)
7. Rescher, N.: The revolt against process. *J. Philosophy* 59 (1962) 410-417
8. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE. Proceedings of EKAW 2002, Sigüenza, Spain (2002)
9. Bittner, T., Donnelly, M., Smith, B.: Endurants and Perdurants in Directly Depicting Ontologies. AI Communications. IOS Press, Amsterdam (2004)
10. Odum, H.T.: Systems ecology. John Wiley and Sons, New York (1983)
11. Keller, R.M., Dungan, J.L.: Meta-modeling: a knowledge-based approach to facilitating process model construction and reuse. *Eco. Mod.* 119 (1999) 89-116
12. Jarrar, M., Demy, J., Meersman, R.: On Using Conceptual Data Modeling for Ontology Engineering. *J. on Data Semantics* 1(1) (2003) 185-207
13. Ford, E.D.: Scientific methods for ecological research. Cambridge University Press, Cambridge (2000)
14. Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social Roles and their Descriptions. Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004), Whistler, Canada (2004)
15. Klein, M., Noy, N.F.: A Component-Based Framework for Ontology Evolution. Workshop on Ontologies and Distributed Systems at IJCAI-2003, Acapulco, Mexico (2003)
16. Todorovski, L., Dzeroski, S.: Using domain knowledge on population dynamics modeling for equation discovery. Proceedings of the Twelfth European Conference on Machine Learning. Lecture Notes in Computer Science, Springer Verlag, Berlin (2001) 478-490
17. Blackford, J.C., Radford, P.J.: A structure and methodology for marine ecosystem modeling. *Neth. J. Sea Res.* 33(3/4) (1995) 247-260
18. Baretta-Bekker, J.G., Baretta, J.W., Koch Rasmussen E.: The microbial food web in the European Regional Seas Ecosystem Model. *Neth. J. Sea Res.* 33(3/4) (1995) 363-379
19. Smith, J.M.: Models in ecology. Cambridge University Press, Cambridge (1979 ed.) (1974)
20. Nihoul, J.C.J.: Modelling marine ecosystems as a discipline in Earth Science. *Earth-Science Reviews* 44(1) (1998) 1-13

Supplementary material: The MicrobialLoop and Pollution ontologies, a color version of Figure 5, and the initial mapping between ML and ontological categories are available online at <http://www.meteck.org/supplDILS.html>.