TOWARDS A DESCRIPTION LOGIC FOR SCIENTIFIC MODELING

Jean-Pierre Müller[1], Hasina Lalaina Rakotonirainy[2] and Dominique Hervé[3]
[1]*GREEN, CIRAD, Campus international de Baillarguet, Montpellier, France*
[2]*ENI, University of Fianarantsoa, Fianarantsoa, Madagascar*
[3]*IRD, MEM, University of Fianarantsoa, Fianarantsoa, Madagascar*

Keywords: Ontology, Metrology, Geographic representation, Knowledge representation.

Abstract: The classical ontologies are based on description logics. Most of the proposed variants fit within the logical framework, with the exception of the introduction of data types. This later extension is suitable for computer scientists but not appropriate for scientists in general. Indeed, most scientists use quantities with full unit systems as defined in metrology. More specifically, the geomaticians use, in addition to attributed or relational structures, what they call coverages, i.e. mappings from coordinates into data. Separate efforts have been made to formalize these aspects but none coped with all of them in an integrated knowledge representation framework. The aim of this paper is to propose description logic extensions able to integrate these various aspects into the general framework of knowledge representation, as a way to talk about matter and space.

1 INTRODUCTION

To model complex systems, (Villa et al., 2009) propose to distinguish three categories of modeling platforms: 1) programming frameworks like Repast (Collier, 2003), 2) declarative modeling environments like Stella (Richmond and Peterson, 2000) and 3) semantic modeling platforms. The later category is further divided into two approaches. The mediation approach where the sub-models inputs and outputs are documented for better integration. The knowledge-driven approach where the model content is itself described using knowledge representation approaches. The intent is "to exploit the formalized semantics of natural systems to unify representations of data and metadata, improve their usability in scientific workflows, and ease the definition of dynamic models" (Villa et al., 2009). In the Mimosa platform, (Muller, 2010) uses ontologies to specify entirely a model, as advocated in (Muller, 2007). The ontologies are then mapped into a simulation model based on DEVS (Zeigler et al., 2000).

The ontology we are using in Mimosa is equivalent to the $\mathcal{ALQ}^{(\mathcal{D})}$ description logic, i.e. with roles, cardinality restrictions and the base data types (integers, doubles, strings). However our experience of using this ontology for complex eco-sociosystems reveals a systematic use of quantities and complex spatial structures. Although the definitions of these quantities and structures are expressible with standard ontologies (see, for example, (Brilhante, 2004)), their systematic use suggests to incorporate them as first class citizens in the formalism as it was made with data types.

The aim of this paper is to propose extensions to description logics. It appears to be a general attempt to semantically incorporate continuous matter (including space and time) where logics are only based on objects. A first section introduces the usual syntax and semantics of description logics. The next section formulates the requirements. Then we propose syntactic extensions to description logics with its associated semantics before concluding.

2 CLASSICAL DESCRIPTION LOGICS

This section recalls the description logics syntax and semantics to define where we are starting from.

2.1 The Syntax

An ontology semantics is formalized with description logics. The most common language called $\mathcal{ALC}$ (for Attributive Language with Complement) is based on the triplet $\mathbb{L} = \langle \mathbb{C}, \mathbb{P}, \mathbb{O} \rangle$, where $\mathbb{C}$ is the set of concept

Müller J., Lalaina Rakotonirainy H. and Hervé D..
TOWARDS A DESCRIPTION LOGIC FOR SCIENTIFIC MODELING.
DOI: 10.5220/0003634201830188
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2011), pages 183-188
ISBN: 978-989-8425-80-5

names, $\mathbb{P}$ the set of role (or relation) names, and $\mathbb{O}$ the set of individual names. The triplet is called the *signature* of the language. Based on this signature, three sets of constructs are defined: the concepts, the terminological axioms and the assertional axioms.

In $\mathcal{ALC}$ the set of possible *concepts* is recursively defined as follows:

- $\top$ is the everything concept;
- $\bot$ is the nothing concept;
- every $C \in \mathbb{C}$ is a concept;
- $\neg C$: the negation of a concept C is a concept;
- $C \sqcap D$: the intersection of two concepts C and D is a concept;
- $C \sqcup D$: the union of two concepts C and D is a concept;
- $\forall r.C$: the universal restriction of a concept C by a role $r \in \mathbb{P}$ is a concept;
- $\exists r.C$: the existential restriction of a concept C by the role $r \in \mathbb{P}$ is a concept.

Intuitively, these constructs allow to derive concepts from other concepts, the last two constructs introducing attribute and/or relation definitions among concepts.

The set T of *terminological axioms* is defined as follows:

- $C \sqsubseteq D$: states that the concept C is included in D;
- $C \doteq D$: when $C \sqsubseteq D$ and $C \sqsupseteq D$, sometimes called a concept definition when $C \in \mathbb{C}$.

The set of terminological axioms forms the TBox or conceptual model. Intuitively, these axioms introduce concept inheritance and definition. The *atomic concepts* are $\top$, $\bot$ and the concepts that do not appear in the left-hand side of the terminological axioms. The other concepts are called the *derived concepts* in the set-theoretical sense.

As an example, we can define an agent, a member or a community, in the following way:

- $Agent \sqsubseteq \forall name.String$: literally, the set of agents is included into the set of everything that has a name of type String.
- $Community \sqsubseteq (\neg Agent \sqcap \forall name.String \sqcap \forall chief.Member)$: the set of communities is included in the set of everything that is not an agent but has a name and a chief.
- $Member \sqsubseteq (Agent \sqcap \forall chief.Member \sqcap \forall group.Community)$: the set of members is included in the set of agents that have a chief and a community.

It is equivalent to descriptions in any frame-like representation language but with a richer expressivity (for example that communities cannot be agents).

The set A of *assertional axioms* is defined as follows:

- $C(a)$: states that an individual $a \in \mathbb{O}$ is an instance of the concept C;
- $r(a,b)$: states that the pair of the individuals $a,b \in \mathbb{O}$ is an instance of the role $r \in \mathbb{P}$.

The set of assertional axioms forms the ABox or concrete model. Intuitively, these axioms describe a concrete system made of categorized individuals and relations.

As an example, we can define a community and a member: $Community(c1)$, $name(c1, "Antontona")$, $Member(p1)$, $name(p1, "Hasina")$, ...

A *knowledge base* is a pair $\langle T, A\rangle$ of axioms, although, most of the time, only T is given. Finally, an *ontology* O is a pair $\langle \mathbb{L}, \langle T, A\rangle\rangle$.

2.2 The Semantics

The above-described language admits a set-theoretic interpretation I which is given by a pair $\langle \Delta, \pi\rangle$ where:

- Δ is a set of objects called the domain of discourse;
- π is a function attributing a meaning to the signature and recursively to concepts in the following way[1]:
  - $\pi(C \in \mathbb{C}) = \{x_i | x_i \in \Delta\}$
  - $\pi(r \in \mathbb{P}) = \{(x_i, y_i) | x_i, y_i \in \Delta\}$
  - $\pi(o \in \mathbb{O}) = x \in \Delta$
  - $\pi(\top) = \Delta$
  - $\pi(\bot) = \emptyset$
  - $\pi(\neg C) = \{x_i | x_i \notin \pi(C)\}$
  - $\pi(C \sqcap D) = \{x_i | x_i \in \pi(C) \wedge x_i \in \pi(D)\}$
  - $\pi(C \sqcup D) = \{x_i | x_i \in \pi(C) \vee x_i \in \pi(D)\}$
  - $\pi(\forall r.C) = \{x_i | \forall y.(x_i, y) \in \pi(r) \supset y \in \pi(C)\}$
  - $\pi(\exists r.C) = \{x_i | \exists y.(x_i, y) \in \pi(r) \wedge y \in \pi(C)\}$

Given these definitions, an interpretation I is a model for the axioms according to the following conditions:

- $I \models C \sqsubseteq D$ if and only if $\forall x, x \in \pi(C) \supset x \in \pi(D)$ (or equivalently $\pi(C) \subseteq \pi(D)$)
- $I \models C \doteq D$ if and only if $\pi(C) = \pi(D)$
- $I \models C(a)$ if and only if $\pi(a) \in \pi(C)$
- $I \models r(a,b)$ if and only if $(\pi(a), \pi(b)) \in \pi(r)$

[1] The semantics specification style complies with classical logics but not with description logics literature!

Accordingly, an interpretation I is a *model* of a knowledge base $\langle T,A\rangle$ ($I \models \langle T,A\rangle$) if and only if $I \models t$ for all $t \in T$ and $I \models a$ for all $a \in A$. If no model exists for a knowledge base, the knowledge base is inconsistent.

2.3 Some Existing Extensions

Other description logics exist, qualified with letters, which define some restrictions or extensions. Existing extensions are, for example:

- $\mathcal{O}$: introduces concepts as sets of individuals;
- $\mathcal{N}$: for cardinality restrictions;
- $\mathcal{Q}$: for fully qualified cardinality restrictions;
- $^{(\mathcal{D})}$: when data types (integer, double, etc.) and values are introduced;

For example, in the $^{(\mathcal{D})}$ description logic, the basic data types (integer, double, etc.) are introduced among the atomic concepts as well as the corresponding data values (45, 10.5, true, etc.) among the atomic individuals. The existence of data values is equivalent to an (almost) infinite set of assertional axioms for all the instances of the basic types.

3 REQUIREMENTS

Our experience in designing large models with scientists of various disciplines (Aubert et al., 2010; Belem et al., 2011) is the following:

- they do not use data types and data values but quantities (i.e. length, weight, etc.) and measures (i.e. values with units);
- the structures are accessed by coordinates and not only (role) names;
- there is a variety of points of view, possibly of the same things.

Philosophically a quantity is a property which exists as a magnitude or a multitude. A physical quantity, as defined by the International Vocabulary of Metrology, 3rd edition, is a property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference. The International System of Quantities defines seven quantities from which all the others can be defined: the length, the time or duration, the mass, the electric current, the thermodynamic temperature, the amount of substance and the luminous intensity. A quantity is measured by a real number and a unit. The international system proposes to measure length in meters (m), duration in seconds (s), mass in kilograms (kg), the electric current in amperes (A), the temperature in degrees Kelvin (K), the amounts of substance in moles (mol) and the luminous intensity in candelas (cd). All the other units can be obtained by combining these units with $*$, $/$ and ˆ and multiplication with some factors (e.g. $1000 * N * cd^2/m$ also named $kilo - N * cd^2/m$). When dividing a unit by itself, the resulting unit is said dimensionless.

All the attributes of an object describe its *qualities* in the philosophical sense. In effect, in the philosophical language, ”being red” or ”having 1.6 meters” are qualities of individuals. Therefore the measures are just descriptions of the physical qualities of individuals which appear to be quantitative. Very often a set of these qualities (height, age, weight, etc.) is necessary. The set acts as a coordinate in the space of physical qualities. Therefore a coordinate is a vector of qualities. We will use this definition in the following.

If a coordinate is a position of an individual in a space of qualities, the space itself is an object where individual objects or descriptions can be obtained given a coordinate. These mappings from coordinates into individuals are very often used in complex system modeling . The geomaticians call them coverages in the particular case where coordinates are only made of lengths or angles.

Finally, the coordinates are measured relative to a reference. If we refer to qualities in general (and not only the physical qualities), even how we name things is relative to a context or a point of view which acts as a terminological reference system. Coordinates are ways of naming things as a terminology is a way to name objects.

4 OUR PROPOSITION

4.1 The Syntax

We propose to define the following atomic concepts instead of the data types:

- *Name* is the concept of all possible strings of characters. We distinguish it from the ”String” data type to keep us apart from any programming notion. However, the corresponding individuals are just strings.
- $\{\ldots,o_i,\ldots\}$ where each o_i is in $\mathbb{O}$, is a concept, called an *enumeration*. The corresponding description logic is therefore of type $\mathcal{O}$. The construct $C \doteq \{\ldots,o_i,\ldots\}$ is both considered a terminological axiom and a set of assertional axioms of the form $C(o_i)$ for each o_i.
- $(\ldots,o_i,\ldots)$ where each o_i is in $\mathbb{O}$, is a concept, called a *series*. In terms of instance, it is similar to

$\{\ldots,o_i,\ldots\}$, but the elements are considered ordered. The construct $C\doteq\{\ldots,o_i,\ldots\}$ is also considered a terminological axiom and a set of assertional axioms of the form $C(o_i)$ for each o_i as well as $<(o_i,o_j)$ for all the appropriate couples.

We also want to introduce two derived concepts:

- $set(C)$ where C is a concept, is also a concept, called a *set*. It is the set of all the sets of elements of $\pi(C)$. Of course sets of sets are possible.
- $range(C,o_1,o_2)$ where C is a concept and o_1 and o_2 are the individual names of elements of $\pi(C)$, is also a concept called a *range*. The syntax could be extended for allowing opened, closed or semi-opened (or semi-closed) intervals. A range is only possible if C is ordered.

Most importantly, we introduce the following constructs for dealing with continuous matter as qualities, coordinates and mappings.

The *quantities* are predefined atomic concepts (e.g. Length, Weight, Duration, etc.) entirely replacing the data types. At least the seven physical quantities mentioned in section 3 must be defined. Additional ones can be provided as needed. The instance of a quantity is a measure. A name of a measure is of the form rU where $r\in\mathbb{R}$ and U is a unit depending on the quantity it is an instance of (e.g. $1kg$, $50.3m$, $12cd$, etc.). The assertional axiom $C(rU)$ is assumed where C is the quantity measured with the unit U (e.g. $Weight(3.2kg)$).

The *coordinate* concepts are derived concepts defined by the construct $\langle C_1,\ldots,C_n\rangle$ where C_i are atomic concepts (e.g. $\langle Weight, Length, \{low, medium, high\}\rangle$) or ranges. We consider $\langle o_1,\ldots,o_n\rangle$ where o_i are individual names as an individual name for a coordinate (e.g. $\langle 1kg, 5.3m, high\rangle$ is a coordinate name). We can have coordinates over unbounded spaces by having at least one concept C_i denoting an unbounded set (e.g. Weight). A coordinate over a bounded space can be specified either by having each C_i denoting a bounded set (e.g. $range(Weight,0kg,100kg)$) or by defining a range on a coordinate concept (e.g. $range(\langle Weight, Length, \{low, medium, high\}\rangle$, $\langle 0kg, 0m, low\rangle, \langle 100kg, 10m, high\rangle))$. Therefore, we consider a range over a coordinate concept as a coordinate concept.

For dealing with indexed spaces, we propose to extend the set of roles $\mathbb{P}$ with the coordinates. Therefore we propose to introduce the expressions: $\forall R.C$ and $\exists R.C$ where both R is a coordinate concept and C is a concept. Therefore we can define concepts as $Elevation \sqsubseteq \forall\langle Length, Length\rangle.Length$, i.e. as a two-dimensional map. Similarly, a space can be defined as a set of named places: $Space \sqsubseteq \forall\langle Name\rangle.Place$. This extension is the most important one, introducing a limited form of second-order quantification for tractability.

To take into account the multiplicity of points of view, one step is to introduce a set of indexed ontologies O_i where $i\in I$ and a notation $i:C$ for any construct C. The later notation allows to reference the construct as described in ontology i. If we want a real modularity, each ontology O_i has his own interpretation $\langle\Delta_i,\pi_i\rangle$ (see for example (Jie Bao and Honavar, 2006)) otherwise a single interpretation for all O_i is enough. Consequently, a number of new axioms must be introduced to build bridges between the various ontologies expressing the points of view. We will not further explore this issue in this paper.

4.2 The Semantics

To express the semantics of the proposed constructs, we have to extend slightly the interpretation $I=\langle\Delta,\pi\rangle$. Δ must include the measures (i.e. a couple (r,u) where $r\in\mathcal{R}$ and u is a unit), and the strings. π is extended as follows:

- $\pi(Name)=\{x_i|x_i\in String\}$
- $\pi(\{\ldots,o_i,\ldots\})=\{\pi(o_i)|o_i\in\{\ldots,o_i,\ldots\}\}$
- $\pi((\ldots,o_i,\ldots))=\{\pi(o_i)|o_i\in\{\ldots,o_i,\ldots\}\}$, and for each o_i,o_j such that $i<j$, $\pi(o_i)<\pi(o_j)$
- $\pi(set(C))=2^{\pi(C)}$
- $\pi(range(C,o_1,o_2))=\{x_i|x_i\in\pi(C)\wedge\pi(o_1)\leqslant x_i\leqslant\pi(o_2)\}$
- $\pi(rU)=(r,U)$ where $r\in\mathcal{R}$ and U is a unit
- $\pi(\langle\ldots,C_i,\ldots\rangle)=\{(\ldots,x_i,\ldots)|\forall i,x_i\in\pi(C_i)\}$
- $\pi(\forall R.C)=\{x_i|\forall y,r_i\in\pi(R).(x_i,y)\in\pi(r_i)\supset y\in\pi(C)\}$
- $\pi(\exists R.C)=\{x_i|\exists y,r_i\in\pi(R).(x_i,y)\in\pi(r_i)\wedge y\in\pi(C)\}$

The resulting semantics is relatively straightforward and does not introduce anything which does not already exist in the classical semantics but the strings and measures as distinguished individuals within Δ. A noticeable exception is the introduction of a second order construct.

4.3 Discussion

As a consequence of the new concept constructs, we extend the set of individuals $\mathbb{O}$ with particular names: rU where $r\in\mathbb{R}$ and U is a unit depending on the quantity it is an instance of. This notation can be easily extended to the colors because colors are well standardized now, as well as the currencies using the norm ISO 4217, or the dates.

It remains to explore what it means for concepts to have roles. We say that a concept C has a role r if we have $C \sqsubseteq \forall r.D$ or $C \sqsubseteq \exists r.D$ in the terminological axioms. In the expressions $\forall r.D$ and $\exists r.D$, we say that the role is of type D. Usually, the roles of a concept are partitioned into two sets: the *attributes* and the *relations*. The attributes are the role of which type is a data type (that we do not use). The relations are all the other roles. Semantically, we identified the following distinctions: 1) the *attributes* are the roles of which type are qualities, 2) the *relations* are the roles that describe topological relations in a broad sense. It can be geometrical, social or temporal, 3) the *mappings* are the roles that give access to a coverage in the geographical sense of a mapping from individuals into individuals that are all of the same type. A concept only with attributes is called a *simple concept*. It corresponds to the notion of simple feature in OpenGIS and can be mapped very naturally with a database schema. The relations define semantic graphs. The mappings can be implemented using a generalized form of coverages.

Semantically, it is assumed that mappings are defined relative to various reference systems. The specifications of OpenGIS are using such reference systems for dealing with coordinates in the huge variety of projection systems (UTM, WSG, etc.). The use of the ontology indexed notations in modular ontologies suggests the possibility to unify the concept of local ontology with the concept of reference system. This track is being pursued but will not be further elaborated in this paper.

5 IMPLEMENTATION

In this section, we shortly describe the chosen implementation of the ontologies as formalized by the proposed extension of description logics. For implementing the concepts (see figure 1), we make the distinction between the quality concepts, the coordinate concepts and all the others (simply called concepts). The coordinates are vectors of qualities. Moreover, the next step is to make them relative to a reference system, while the qualities are absolute.

Regarding the quantities, we have fully implemented the unit definition mechanisms as described by the International System of Units, as well as the possibility to define all the possible quantities. This implementation is inspired from the jsr-275 attempt (JScience, 2009). However jsr-275 is defined for compile time use of quantities and measures. In particular, the defined quantities are subclasses of the Quantity class, making the introduction of new quantities difficult. Moreover the access and use of the list of defined quantities at execution time is impossible. Consequently, we defined our quantities as instances of the Quantity class, making it declarative and easily extensible.

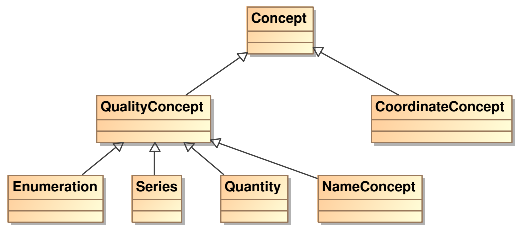

Figure 1: The concept classes.

Notice that the derived concepts are not defined in a separate class because we have chosen to represent the derivations as relations among concepts. As a consequence, figure 2 shows all the derivations we have included in our description logic; namely the union, intersection, complement, inclusion and roles as in standard description logics, but also the range, set and mapping.

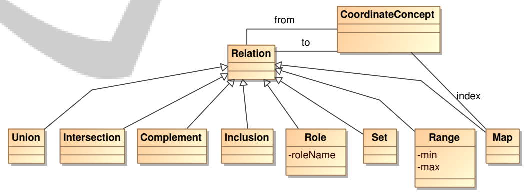

Figure 2: The relations among concepts.

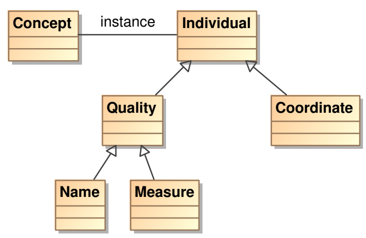

Figure 3: The individual classes.

The implementation of the individuals reflects the particular roles some of the instances have (figure 3). In particular, the strings, measures and coordinates are distinguished. Otherwise, as for the concepts, the relationships among individuals are implemented as relations implementing the various links (from role names to individuals and from coordinates to individuals). In the real implementation, the quality class does not exist because an individual is a quality if it is an instance of a quality concept.

Additionally, the ontologies introduce name spaces where the names are linked to the concepts for the names in $\mathbb{C}$, to the individuals for the names in $\mathbb{O}$ and to the roles for the names in $\mathbb{P}$.

6 CONCLUSIONS

In this paper, we have argued that real world modeling with scientists from various disciplines does not accommodate the use of pure mathematical or programming notions like the data types. In particular, they need to describe the quantities they measure in the real world using units. Beyond using measures, the world they are dealing with is not only made of objects but also of matter and spaces, which, most of the time, are continuous, bounded or unbounded entities. Although a semantics of sets, as we have shown, can accommodate continuity (with continuous sets) and boundedness (by introducing order and sets as intervals), there is a need to incorporate the proper constructs as first class citizens for better expressiveness: i.e. the quantities, the coordinates and the mappings. This paper has proposed such constructs with the associated semantics. This proposition, as well as partly what follows as a perspective, has been implemented as an extension to Mimosa ((Muller, 2010), http://mimosa.sourceforge.net/).

The immediate perspective is to introduce the reference systems. In effect, a coordinate is not absolute but is always relative to a reference system. If two coordinates are given in two different reference systems, they must be mapped from one into the other. OpenGIS has defined the mechanisms for doing so among geographic coordinates, but these mechanisms should be extended. Not so surprisingly, in multidisciplinary contexts, a terminology is relative to who is talking as well. Two names in different ontologies must be mapped from one into the other. Bridge rules are the mechanisms for doing so as described in (Jie Bao and Honavar, 2006). What precedes suggests a possibility to unify this problem of mapping a multiplicity of reference systems including the ontologies. The next step is to extend the set of concept relations with bridge rules in order to fully implement modular ontologies.

Another ongoing work is to formulate the Mirana conceptual model (Aubert et al., 2010) we are currently working on using the proposed extension. This would illustrate the expressivity of the proposed description logic.

ACKNOWLEDGEMENTS

This work has been jointly financed by IRD and CIRAD.

REFERENCES

Aubert, S., Muller, J.-P., and Ralihalizara, J. (2010). MIRANA: a socio-ecological model for assessing sustainability of community-based regulations. In *International Congress on Environmental Modelling and Software Modelling for Environment's Sake*, pages 1–8, Ottawa, Canada.

Belem, M., Bousquet, F., Muller, J.-P., Bazile, D., and Coulibaly, H. (2011). A participatory modeling method for multi-points of view description of a system from scientist's perceptions: application in seed systems modeling in Mali and Chile. In *ESSA 2011, submitted*, Montpellier.

Brilhante, V. (2004). An Ontology for Quantities in Ecology. In Hutchison, D. and al., editors, *SBIA 2004*, pages 144–153, Berlin, Heidelberg. Springer Berlin Heidelberg.

Collier, N. (2003). Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research*.

Jie Bao, D. C. and Honavar, V. G. (2006). Modular Ontologies - A Formal Investigation of Semantics and Expressivity. In *The Semantic Web – ASWC 2006*, pages 1–16.

JScience (2009). Jscience. http://jscience.org/jsr-275/api/.

Muller, J.-P. (2007). Mimosa: using ontologies for modeling and simulation. In *Advanced Semantics Technologies*, pages 1–5, Bremen, Germany.

Muller, J.-P. (2010). A framework for integrated modeling using a knowledge-driven approach. In *International Congress on Environmental Modelling and Software*, pages 1–8, Ottawa, Canada.

Richmond, B. and Peterson, S. (2000). *STELLA: An Introduction to Systems Thinking*. High Performance Systems Inc.

Villa, F., Athanasiadis, I. N., and Rizzoli, A. E. (2009). Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. *Environmental Modelling & Software*, 24:577–587.

Zeigler, B. P., Praehofer, H., and Kim, T. G. (2000). *Theory of modeling & simulation, integrating discrete event & continuous complex dynamic systems (2nd Ed.)*. Academic Press, New York.