Semantic model integration: an application for OWL

Gary Polhill and Nick Gotts

The James Hutton Institute, Craigiebuckler, Aberdeen. AB15 8QH
{gary.polhill, nick.gotts}@hutton.ac.uk

Abstract. Many agent-based models, especially those embedded in environmental systems are formed of coupled submodels each providing a component of the whole model. Coupled socio-environmental modelling is not trivial, with many authors reporting a need to ensure that the resulting model has a consistent, integrated ontology. The matter can be expected to be no different in the specific case of coupling agent-based and environmental models. We argue that representing the state and structure of the simulation at each time step using OWL ontologies can at least provide *some* automated consistency checks for a proposed coupling that would not be possible using black-box model coupling frameworks. We further argue that *algorithmic conflicts* in coupled modelling exercises are an additional problem for semantic integration of models in comparison with semantic integration of data.

Keywords: Integrated modelling, semantic integration, OWL ontologies.

Model integration as a problem of semantic heterogeneity

The problem of *semantic heterogeneity* arises from the integration of distributed heterogeneous databases – for example, when companies merge and their personnel databases need to be integrated. Various forms of semantic heterogeneity exist, including naming conflicts (the same name is used for different entities; or different names for the same entity), scaling conflicts (where concepts are represented at different spatial or temporal scales), confounding conflicts (where concepts appear to have the same meaning, but don't), and representation conflicts (where concepts are represented in different ways) [1].

There is no reason in principle why, when coupling submodels together, similar problems should not occur. 'Black box' approaches to model coupling, such as the popular Open Model Interface Environment (OpenMI [2, 3]) framework are thus potentially vulnerable: If a coupled model is viewed as a series of submodels, each of which has inputs and outputs that are connected to other submodels, black box frameworks provide no safety checks that the coupled whole is 'ontologically consistent' – a matter that has been recognised as important in other areas attempting model coupling (e.g. [4, 5]). Black box frameworks can provide a beguiling sense of legitimacy to the coupled whole, and have the added advantage that legacy models do not need reimplementing (they can be executed with some standards-conforming wrapper code), and intellectual property rights over their contents can be protected. (Not everyone, it seems, is concerned about software licensing to promote good science [6].)

In coupled *modelling* environments (as opposed to data integration), however, there is a bigger problem than ontological consistency, since models add dynamics. These dynamics change the assertions that specify the model's state from one time slice to another. This creates the potential for what could be termed *algorithmic conflicts*. To understand the issue, consider two submodels A and B, with different (possibly overlapping) sets of input and output variables. Each submodel makes a series of computations on its input variables to produce the output variables. However, suppose there is a conceptual link between an intermediate variable in submodel A, and another in submodel B (Fig. 1) – they correspond to the same entities or properties of the real world. These computations are hidden in black box environments, so there is no way even to know that it happens. Further, because in social and environmental models there is often more than one way to compute a particular variable, there is no guarantee that the conceptually linked variables in submodels A and B will correspond with each other. We can apply Wilensky and Rand's [7] article outlining the different degrees of similarity between model behaviour to the conceptually linked variables: Ideally, we would want the conceptually linked variables in A and B to have the same value. We might be able to living with a statistically significant approximate similarity. However, it is possible that the values of the conceptually linked variables will not even be qualitatively similar (e.g. they both increase and decrease in value under the same circumstances). At least, there is no theoretical guarantee of their similarity.

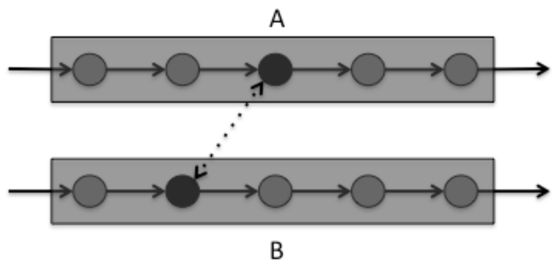

Fig. 1. Two submodels A and B, with conceptually linked intermediate variables (dotted line).

Applying OWL ontologies

We have argued separately [8] that using OWL ontologies to represent the state and structure of a model at each time step improves transparency. This transparency can also help with model integration. Suppose submodel A is considered as a function mapping some subset of the whole model's state at time T to that at time T + 1. Submodel A is implemented as a parametric polymorphic algorithm: all variables used by the algorithm, including intermediate variables, are implemented simply as 'empty' names in submodel A's ontology – to get a value, these names need to be declared equivalent to properties in the integrated model ontology. The same applies to submodel B. Fig. 2 illustrates the idea: the conceptually linked intermediate variables in submodels A and B are exposed (along with all the other intermediate variables) and declared equivalent to properties (which may have other ontological relations among them) in an OWL ontology Ω that provides an integrated perspective on the model as a whole. Automated reasoning services can now be used to detect where there is an inconsistency (e.g. if the intermediate variables in A and B have different datatypes, or are relations that refer to disjoint classes). Semantic integration is an area of active on-going research, with many challenges still to overcome [9]: hence the proposed approach cannot *guarantee* a proposed coupling is free from inconsistencies. It at least provides some checks.

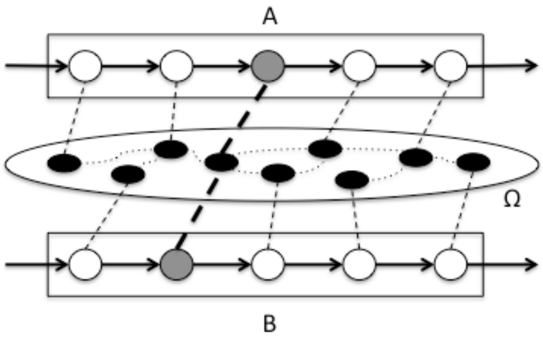

Fig. 2. Two submodels A and B, with their intermediate variables as empty slots that are populated from an integrated perspective of the model provided by OWL ontology Ω.

Example

For a hypothetical example, we consider the integration of FEARLUS [10, 11] with a biodiversity model, SPOMM, to model the impact of agricultural land use decisions on biodiversity and the effectiveness of agri-environmental incentive schemes in reducing it [12]. SPOMM stands for Stochastic Patch Occupancy Metacommunity Model, and is based on Moilanen's [13] SPOMSIM model. The latter is a *metapopulation* model, which means it models the spatial distribution of a single species. SPOMM, however, is a *metacommunity* model because it models the spatial distribution and interaction of multiple species.

Briefly, FEARLUS models farmers making land management decisions on the parcels of land they own. Optionally, a government agent may operate to issue financial incentives that meet with its policy. When coupled with SPOMM, the land management decisions are mapped onto habitat availability, which the SPOMM then uses to calculate occupancy for each species. A government agent may then monitor the presence of species and issue incentives accordingly.

Imagine that FEARLUS, SPOMM and SPOMSIM have been implemented using OWL ontologies to represent the state and structure of the model and any time. For the purposes of this example, in the case of SPOMSIM, a functional object property is used to relate a *Patch* to the *Species* occupying it to indicate that in this metapopulation model a *Patch* only ever has zero or one occupying *Species*, whereas in the SPOMM, a non-functional property is used to indicate that in this metacommunity model a *Patch* has zero or more occupying *Species*. When the government agent in FEARLUS wants to count the species on a *Patch* (assuming it were using species richness or α-diversity as its biodiversity metric), it would look for a non-functional property relating the *Patch* to the *Species* occupying it. This would create an automatically detectable inconsistency when coupling FEARLUS with SPOMSIM that would not occur when coupling with SPOMM. One would then know that coupling such a configuration of FEARLUS with SPOMM is safer than coupling it with SPOMSIM.

Discussion

The use of OWL ontologies in agent-based modelling and related areas is growing. Christley et al. [14] describe an OWL ontology of agent-based modelling approaches, and argue that such ontologies can assist with exposing hidden underlying assumptions in models, among other things. Parker et al. [15] developed the MR POTATOHEAD framework to capture the components that might be expected to appear in an agent-based model of land use and cover change, and implemented it in OWL [16].

Ontologies in other formalisms besides OWL are also being used. Müller et al. [17] use ontologies in the initial stages of model development to describe a conceptual model with stakeholders, which are then used to develop the UML diagrams from which the object-oriented simulation model is eventually coded. Other relevant work includes Villa's [18] model integration architecture, which uses XML to describe the modular integrated components. Bian and Hu [19] emphasise the use of a standard ontology in a discipline to facilitate model interoperability.

There are also links with SysML[1], which is an extension of the Unified Modelling Language diagramming formalism for software systems to allow it to allow more general representations for systems engineering. SysML breaks a system down into modular 'blocks', which are supposed to be reusable. SysML has rules built in to the language that are designed to facilitate model consistency checking, for example, checking consistent units and datatypes at interfaces between properties [20]. Whilst SysML documents a model rather than executing it, well-formed SysML diagrams that conform to the language's semantics can be used to generate code. To provide the same consistency-checking power as the proposed approach using OWL ontologies here, however, each step of the running simulation from the generated code would need to be represented in SysML and consistency-checked.

[1] http://www.sysml.org/

Acknowledgements

This work was funded by the Scottish Government, by the Economic and Social Research Council grant reference RES-149-25-1075 (PolicyGrid II) and by the EU Framework Programme 7 project reference 265155 (LOCAW).

References

1. Bellatreche, L., Xuan Dung, N., Pierra, G. and Hondjack, D. (2006)Contribution of ontology-based data modeling to automatic integration of electronic catalogues within engineering databases. *Comput. Ind.* 57:711-724.
2. Moore, R. V. and Tindall, I. (2005) An overview of the Open Modelling Interface and Environment (the OpenMI). *Environ. Sci. Policy* 8:279-286.
3. Gregersen, J. B., Gijsbers, P. J. A., Westen, S. J. P. and Blind, M. (2005) OpenMI The essential concepts and their implications for legacy software. *Adv. Geosci.* 4:37-44.
4. Frysinger, S. (2002) Integrative environmental modelling. In Clarke, K. C., Parks, B. O., Crate, M. P. (eds.) *Geographic Information Systems and Environmental Modeling*. Prentice Hall Upper Saddle River, NJ, pp. 221-222.
5. Leavesley, G. H., Markstrom, S. L., Restrepo, P. J. and Viger, R. J. (2002) modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling. *Hydrol. Process.* 16:173-187.
6. Polhill, J. G. and Edmonds B. (2007) Open access for social simulation. *JASSS* 10(3):10.
7. Wilensky, U. and Rand, W. (2007) Making models match: replicating an agent-based model. *JASSS* 10(4):2.
8. Polhill, J. G. and Gotts, N. M. (2009) Ontologies for transparent integrated human-natural system modelling. *Landscape Ecol.* 24:1255-1267.
9. Noy, N. F., Doan, A. and Halevy, A. Y. (2005) Semantic integration. *AI Mag.* 26:7-9.
10. Gotts, N. M. and Polhill, J. G. (2009) When and how to imitate your neighours: Lessons from and for FEARLUS. *JASSS* 12(3):2.
11. Polhill, J. G., Sutherland, L.-A. and Gotts, N. M. (2010) Using qualitative evidence to enhance an agent-based modelling system for studying land use change. *JASSS* 13(2):10.
12. Gimona, A. and Polhill, J. G. (2011) Exploring robustness of biodiversity policy with a coupled metacommunity and agent-based model. *J. Land Use Sci.* 6:175-193.
13. Moilanen, A. (2004) SPOMSIM: Software for stochastic patch occupancy models of metapopulation dynamics. *Ecol. Model.* 179:533-550.
14. Christley, S., Xiang, X. and Madey, G. (2004) Ontology for agent-based modeling and simulation. *Proc. Agent 2004, October 7-9, 2004*, pp. 115-125.
15. Parker, D. C., Brown, D. G., Polhill, J. G., Deadman, P. J. and Manson, S. M. (2008) Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: The MR POTATOHEAD framework. In: Lopez Paredes, A. and Hernandez Iglesias, C. (eds.) *Agent-Based Modelling in Natural Resource Management*. pp. 23-51
16. Parker, D. C., Polhill, J. G. and Mussavi Rizi, S. M. (2008) An OWL (Web Ontology Language) representation of the MR POTATOHEAD agent-based land-use change meta-model. *American Association of Geographers Annual Meeting, Boston, MA, April 15-19, 2008*.
17. Müller, J. P. (2007) Mimosa: Using ontologies for modeling and simulation. *Complex07, 8th Asia-Pacific Complex Systems Conference, 2-5 July 2007, Gold Coast, Australia*.
18. Villa, F. (2001) Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. *Ecol. Model.* 137:23-42.

19. Bian, L. and Hu, S. (2007) Identifying components for interoperable process models using concept lattice and semantic reference system. *Int. J. Geog. Info. Syst.* 21:1009-1032.
20. Friedenthal, S., Moore, A. and Steiner, R. (2008) *A Practical Guide to SysML: The Systems Modeling Language*. Burlington, MA: Morgan Kaufmann