2008

A Requirement List for Geo-ontology Tools

Jochen Albrecht
CUNY Hunter College

A Requirement List For Geo-ontology Tools

Jochen ALBRECHT

Abstract

Numerous authors have presented ontology building tools that have all been developed as part of academic projects and that are usually adaptations of more generic tools for geo-spatial applications. While we trust that these tools do their job for the special purpose they have been built for, the GIScience users community is still a long way away from off-the-shelf ontology builders that can be used by GIS project managers.

In our paper, we present a comparative study of ontology building tools described in some twenty peer-reviewed GIScience journal articles. We analyze them from the perspective of two application domains, crime analysis and transportation / land use – both of which are inherently spatial and temporal in nature. For the second example, we developed a database schema, which is sufficiently different from the three main templates for static transportation structures commonly used. For the crime analysis application, we developed a rule base for an agent-based model that had no precursor. In both cases, we were forced to manually code ontologies for use with ESRI-based application software. Based on these experiences, we outline a requirements list of what the tools described in the first part of the paper are missing to make them practical from an applications perspective. The result is an R&D agenda for an important aspect of GIScience.

1 Introduction

Ontologies are en vogue. Many conferences such as GIScience or AGILE are dominated by ontology papers and most relevant journals have had special issues on this topic. Even trade magazines have jumped onto the bandwagon, and it is now considered to be good practice to develop domain-specific data models by paying at least lip service to the underlying ontology. “Traditional” (what is traditional in a discipline that was virtually non-existing ten years ago?) ontology building tools such as OilEd, OntoEdit, PowerLoom, Protégé, or WebODE are a) not geared towards geo-spatial phenomena, which by their very nature are also dynamic, and b) are not exactly aimed at end users such as GIS managers in a regional authority. The situation is similar to the fashion industry, where the models on the runway have little to do with Jo or Jane Public. Even the result is the same, in that we are (all made to) feel bad about our inaptitude to develop sound database and service schemas while we are made to believe that this is what we ought to do.

We now have a generation of peer-reviewed articles that describe ontology building tools for geo-spatial applications and the next section will review them to take stock of what may be regarded as the current state of the art. There is, however, an obvious disconnect between these special purpose built tools and off-the-shelf or preferably public domain ontology builders that can be used by GIS project managers. Section Three describes two projects, where the authors developed complex schemas that on one hand are based on very reasonable needs of the practice, while on the other hand completely beyond the scope of what could be produced by currently existing tools. This suggests a gap in current applied research and we therefore conclude our paper with a list of recommendations on how to address this mismatch.

2 State of the Art – in Theory

General purpose ontology builders and editors have matured, at least from an information scientist's perspective (Duineveld *et al.* 2000, Cardoso & Escorcio 2007). Within academia and in the open source world, Protégé (2007) has become the tool of choice for close to 100,000 registered users, who (in addition to it being free) appreciate its many extensions for different paradigms (frames vs. OWL), multiple APIs, many export formats, and several different editing tools. Originally developed for the medical community, it is now widely used, even in geo-spatial communities (Souza et al. 2006, Lüscher et al. 2007). However, Protégé has its limitations. As Lüscher et al. (2007, p. 9) point out, "Protégé may be too complex for domain experts". They therefore employed a special user interface for creating spatial patterns. Another issue is that while Protégé is well suited for the development of Semantic Web applications, most GIS are still desktop- or heavy server-based, i.e., they require a database schema. Protégé can read database schemas but cannot export to them (database schema may however be created through a JDBC plug-in).

The solution to this is therefore to move to another public domain ontology editor called WebODE (2003). As the name suggests, WebODE is the web-based front-end to what the authors call an Ontology Development Environment (ODE). In its latest incarnation as ODE Semantic Web Services designer (Gómez Pérez *et al.* 2004), this tool limits itself to the creation of OWL-S services, albeit in a very comfortable manner. Of higher value for our purposes here is the still available version of WebODE 2.0, which exports to wide range of data exchange standards such as XML, RDF, OWL, Java/Jess, and UML. The latter two can be used to then create database schemas in, for example, Oracle or MySQL (Corcho et al. 2002). This gets us a long way towards linking original geo-spatial ontology development with the creation of professional GIS database schemata. However, a GIS is more than just a data repository.

As Noy & Guinness (2001) point out, programmers make design decisions based on the operational properties of a class such as methods or procedures, whereas an ontology designer makes these decisions based on the structural properties of a class (type and interface definitions). In consequence, we are still short of ontologies of geographic processes, and ontologies are much easier translated into a database schema than into process models.

Both Protégé and WebODE allow for OWL-S output, i.e., the specification of services, which according to the WorldWideWeb Consortium (W3C 2004) can in turn be modeled as processes. We have, however, yet to see the development of geographic process libraries, notwithstanding the efforts of NASA (Raskin 2004), a few intrepid companies (GOL 2004), and the odd GIScientist (Reitsma & Albrecht 2005, Peachavanish & Karimi 2007), the formal specification of geographic processes is at this point restricted to web services and queries á la Google Maps. This should come as no surprise because the ontology community has not really gone beyond the SNAP/SPAN view introduced by (Grenon & Smith

2003). We are still missing a practical ontology of process that is both proven to be formally correct and at the same time well enough developed to reach to the level of real world applications.

The best we have at this point is UML output based on OWL-S encoded ontologies. As mentioned above, (web) services, the usual target of OWL-S ontologies are a far cry from the complexities of desktop-based spatial decision support systems. The best link that these authors have found so far, is a commercial ontology editor called SemTalk (2008). Originally designed for business workflows, it is an OWL interpreter developed as an extension to MS Visio. Formal consistency is assured with the open source OWL reasoner PELLET (2008). Pre-existing ontologies such as SWEET (Raskin 2004) can be imported and form the foundation for subsequent user-defined extensions. One of the advantages of this tool is its relationship with MS Visio, a wide-spread software package that is also commonly used to define the data models for various ESRI communities (ESRI 2008). This allows for using one and the same tool to define both geo-spatial data structures and processes.

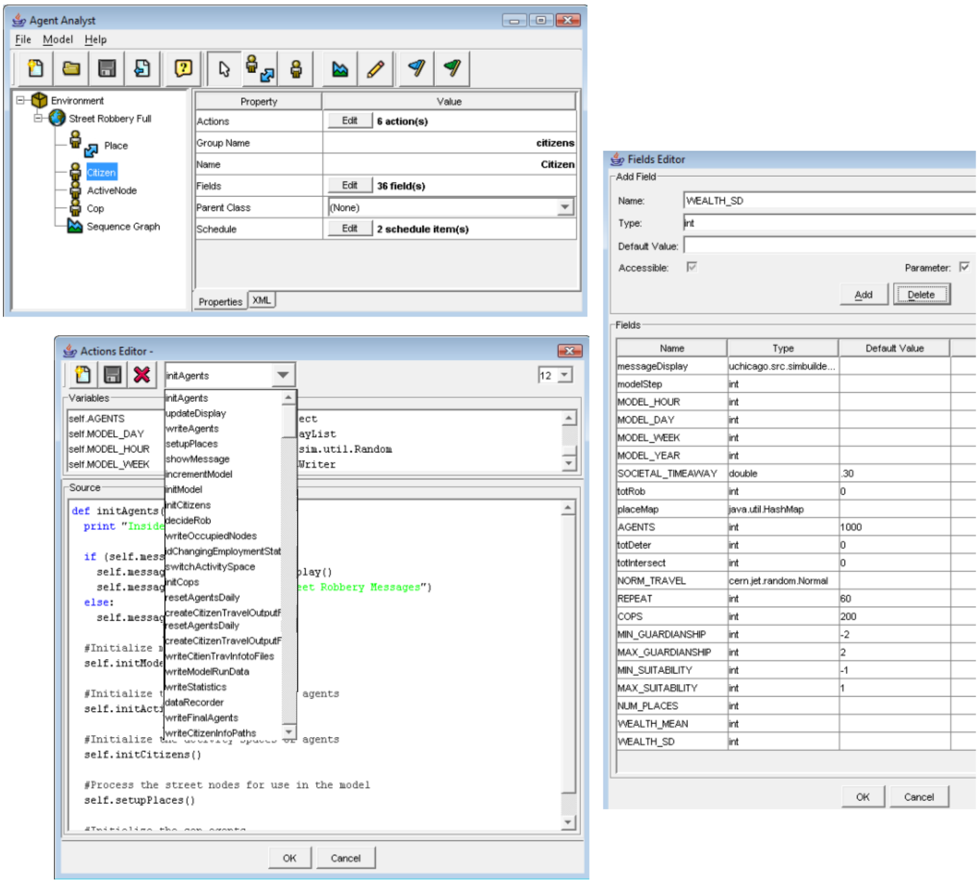
In the next section, we will discuss two larger application examples, where the authors struggled with the definition of new data and process models. This discussion is intended to illustrate the needs from an applied perspective, in this case two university / public agency partnerships.

3 Practical Experiences Highlighting Gaps between Theory and Practice

This section consists of two parts. In the first, we describe and analyze the needs of an agent-based GIS model developed by Groff and described across a range of publications (Groff 2007a, b; 2008). The emphasis in this discussion is on the formal specification of agent behavior in a geographic context. In the second part, we look at a transportation / land use model developed for a regional planning authority. In both cases, the authors are forced to manually code their ontologies rather than using any of the above mentioned tools.

3.1 Formal specification of actions in an agent-based GIS

This model has been described in detail in Groff (2007a, b). We will here instead focus on the process of agent specification, which is documented in form of a sample application of AgentAnalyst (2008). The agent-based component is a Repast (North *et al.* 2006) application that links to ArcGIS using a Python interface. Following the logic of Kuhn (2001) Groff develops a formal framework of agent actions within a well-defined context, in this case crime analysis. Theodorakis *et al.* (1999) and subsequently Cai (2007) describe such a process in abstract terms but show no attempt of an implementation. Groff (2008) develops her specifications the way a programmer would: defining in a painstaking iterative process all the variables and actions of the agent based model (**Fig 1**) and arranging them in the complex schedule that operates at four different hierarchical levels.



Property	Value
Actions	6 action(s)
Group Name	citizens
Name	Citizen
Fields	36 field(s)
Parent Class	(None)
Schedule	2 schedule item(s)

self.AGENTS
self.MODEL_DAY
self.MODEL_HOUR
self.MODEL_WEEK

```
ect
ayList
sim.util.Random
Writer
```

```
def initAgents(
  print "Inside

  if (self.mess
    self.messag                  play()
    self.messag                  eet Robbery Messages")
  else:
    self.messag

  #Initialize m
  self.initMode

  #Initialize t                  agents
  self.initActi

  #Initialize t                  agents
  self.initCitizens()

  #Process the street nodes for use in the model
  self.setupPlaces()
```

Name: WEALTH_SD
Type: int
Default Value:

Name	Type	Default Value
messageDisplay	uchicago.src.simbuilde...	
modelStep	int	
MODEL_HOUR	int	
MODEL_DAY	int	
MODEL_WEEK	int	
MODEL_YEAR	int	
SOCIETAL_TIMEAWAY	double	.30
totRob	int	0
placeMap	java.util.HashMap	
AGENTS	int	1000
totDeter	int	0
totIntersect	int	0
NORM_TRAVEL	cern.jet.random.Normal	
REPEAT	int	60
COPS	int	200
MIN_GUARDIANSHIP	int	-2
MAX_GUARDIANSHIP	int	2
MIN_SUITABILITY	int	-1
MAX_SUITABILITY	int	1
NUM_PLACES	int	
WEALTH_MEAN	int	
WEALTH_SD	int	

Fig. 1: Top-level view of the variables and activities of a street robbery model. The Edit buttons in the top left window open the respective Actions and Fields (variable) editor windows displayed on the bottom and right. The Actions Editor in turn creates the stubs for Java classes and methods.

The total model consists of four types of agents (places, citizens, cops, and activity nodes) that perform 35 defined activities based on the state of 64 variables. It took Groff a good year of her PhD to develop this model. There are no tools that assist in the setup of the model logic, or that would help with the checking for formal consistency of the model. In an agent-based model, the agents are supposed to act depending on context. From a developer's perspective, this context, however, gets fast out of hand because of the combinatorial explosion of possible situations among the many dimensions that Groff's model spans. Ali *et al*. (2007) address this problem with a spatial variant of an online analytical process (OLAP) to keep track of the complexities in the real-time analysis of their multi-agent geosimulation software decision making process. This takes care of some of the implementation issues but leaves the question of ontology-based design tools that this article is about unresolved.

3.2 Formal specification of a land use-based transportation model

The second model to be discussed here has not yet been released to the general public and the description is purposefully general as to not identify the agencies involved. Goal of this spatial decision support system (SDSS) is to model existing flow in a multi-modal network and to find optimal locations for transit hub based on existing parcel-level land use. The resulting geodatabase (all vector data) is approximately 2 GBytes in size, and the procedures of the SDSS combine many hundred GIS operations for a single model iteration. A superficial view of the geoprocessing model is depicted in **Fig 2**, while **Fig 3** provides a top view of the underlying database schema.

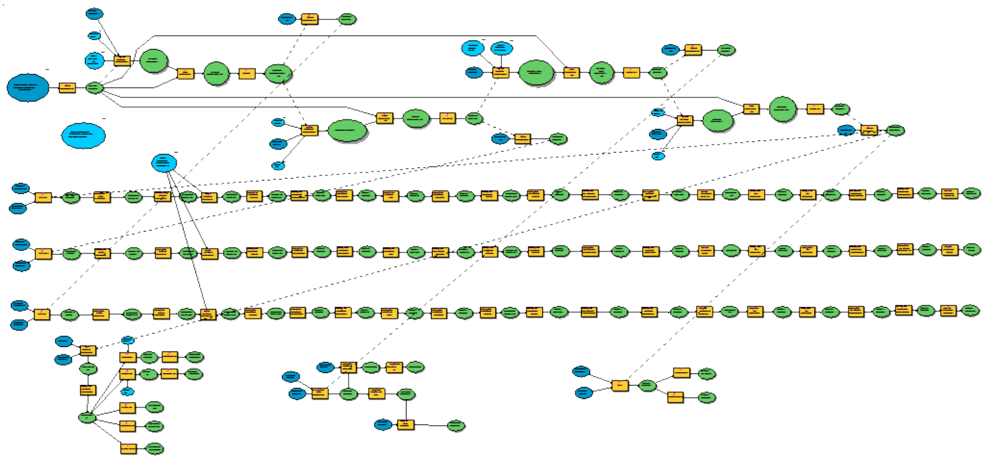


Fig. 2: Top-level view of a land use-based transportation hub siting model.

GIS databases are often treated as mere repositories that once created are frozen in time and form a stable basis for subsequent analysis. The efforts of creating a database are hence neglected in academic literature (notwithstanding publications on management (Huxhold & Levinsohn 1994, Tomlinson 2003, Watson 2005, Obermeyer & Pinto 2007, Yeung & Hall 2007), metadata (Williamson *et al.* 2003, Nogueras-Iso *et al.* 2005, Wootton 2007) or spatial decision support (Geertman & Stillwell 2002)).

Similar to the crime model of the previous section although not at the fine-grained individual level of perpetrators and victims, our very model structure is dynamically changing as we run the model. For instance, the choice of a commuter’s mode or route is influenced by how many other commuters crowd one’s first or second option. Many train riders in Seattle, for example, are familiar with the added number of passengers when flooding closed a highway during exceptional rainstorm in fall of 2007. The same happens on a smaller but cumulatively as effective scale every day in metropolitan areas around the world. From an abstract perspective, the phenomenon is well studied and has become a popular example in complexity theory (Qiu *et al.* 2000, Sheng *et al.* 2005; Selçuk 2007). As a matter of fact, it is one of the hallmarks of complexity theory (Gimblett 2002) that its object arises from innocent looking simple setups. With the exception of a draft paper by Line (2004), there is no ontology of complex systems and we are left with basic tools such as the Unified Modeling Language (Tsang *et al.* 2005), which in itself is not a formal reasoned and in practice lacks the mechanisms to deal with complex systems.

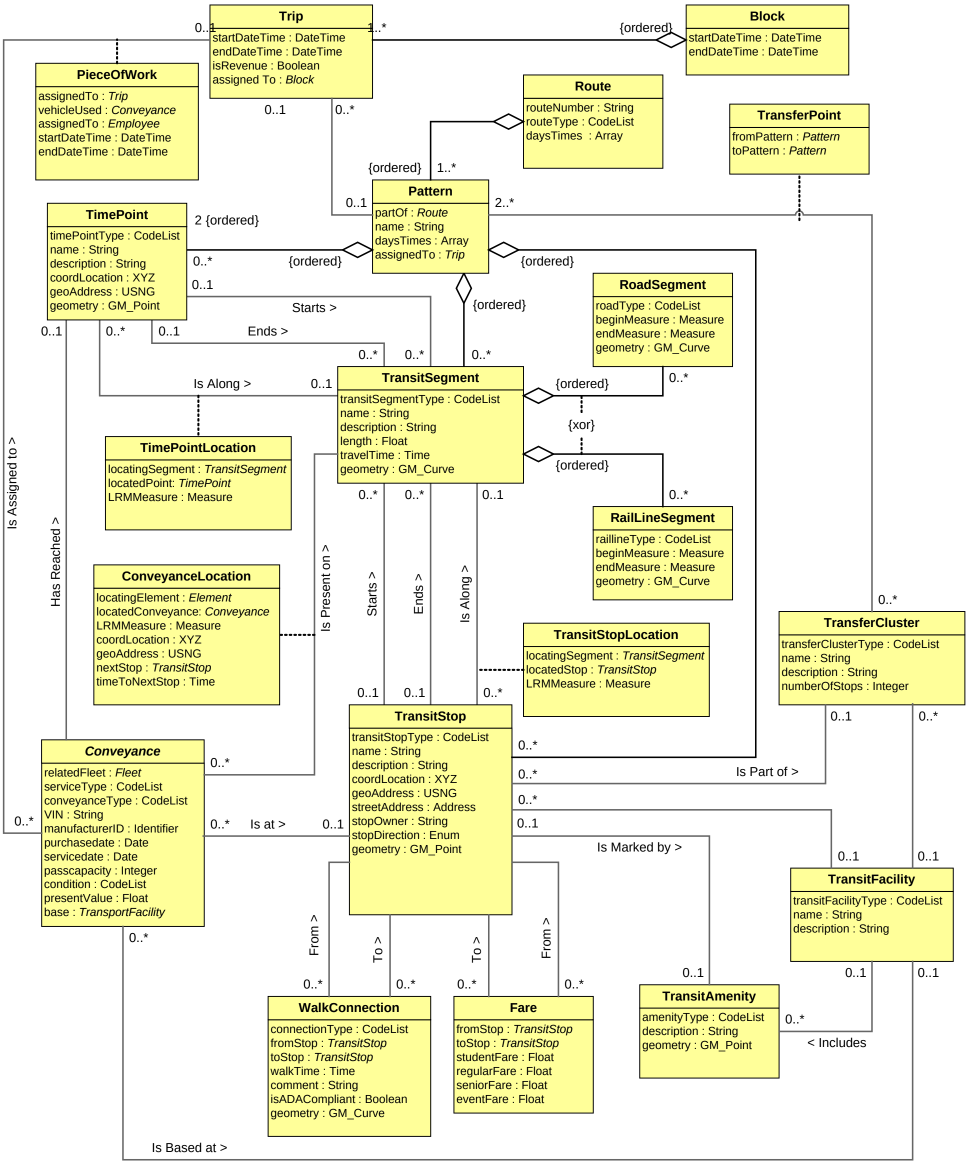


Fig. 3: Top-level view of a land use-based transportation hub siting database schema. The full schema has over two hundred feature classes Geographical Information Systems and Dynamic Modeling via Agent Based Systems

4 Conclusions, aka a Call for a Research Agenda

Based on the above examples we outline a requirements list of what the tools described in the first part of the paper are missing to make them practical from an applications perspective. Klien & Probst (2005), as well as daSilva *et al.* (2005) call for research on methods and tools that support application ontology engineering, e.g., by automating the process of creating application ontologies. We cannot agree more. So, how do we get there? Apparently, the problem is non-trivial; the authors are not aware of any solutions in the wider non-spatial realm.

In a first step, we need to increase our efforts on the low-level ontology side. Grenon & Smith's SNAP/SPAN dichotomy is a neat logical device to compartmentalize views of dynamic phenomena but we need a unified view that allows for the application of one methodology rather than the continued parallel use of two. Reitsma (2005) introduces a radical approach, which does not seem to have found too big a following, yet she seems to have been onto something: Line (2004) and Raskin (2004) independently came up with similar notions of 'flux' and 'trajectory' that do away with the idea that we can capture processes in form of discrete automata steps. In information science, the solution to the latter was introduced by Petri (1962). Petri nets, also known as place / transition nets allow for formally complete specifications, are uniquely suited for dynamic phenomena, and could conceivably be linked with network- as well as with Voronoi-based GIS (e.g. Gold since 1987).The International Standards Organization (ISO/IEC JTC1 2005) is about to accept a Petri Net markup language that will go a long way towards public access to what used to be a rather theoretical tool.

Second, we need spatio-temporal schema editors that assist with the development of complex scheduling and update routines in dynamic models. Even industrial-strength UML tools like Rational Rose (2003) cannot cope with the spatio-temporal granularity of models such as those described in the previous section. SemTalk (2008), described in section 2, comes closest so far, but is constrained by the limitations of OWL-S. An immediate solution for this problem is not visible, as this kind of optimization of a geo-spatial process model specification is not an inherently academic endeavor. Similar to the development of true three-dimensional GIS, we may even miss the inauguration of such tools because they remain behind the closed doors of the defense and oil industry.

Third, and with kudos to an anonymous reviewer, we need to revisit the ontologies of fields discussion of 1998 (Kemp & Vckovsky) and 1999 (Peuquet et al), which lead to wonderful philosophical insights but was never picked up by the tool building community.

Finally, and here we join the tenor of Klien & Probst (2005), a wide-spread move from static GIS repositories to GIS-based process modeling systems requires the development of reusable libraries of process model specifications, similar to their counterpart in the data model world. This is of course dependent on the availability of standardized tools and methodologies – the steps one and two from above. Until then, the conceptual difficulties of bridging domain-specific models (Albrecht 2007) as well as the sheer effort required to develop real-world dynamic GIS will render them a rare and endangered species.

References

AgentAnalyst 2008. Agent-based Modeling Extension for ArcGIS Users. Web site at http://www.institute.redlands.edu/agentanalyst/, last accessed 01/29/08.

Albrecht J 2007. Dynamic GIS. Wilson, J. and S Fotheringham (Eds) *The Handbook of Geographic Information Science*, pp. 436-446. London: Blackwell.

Ali W, Moulin B, Bédard Y, Proulx M and S Rivest 2007. Coupling multiagent geosimulation and spatial OLAP for better geospatial data analysis. *URISA Journal* **19**(2): 23-32.

Cai G 2007. Contextualization of Geospatial Database Semantics for Human-GIS Interaction. *GeoInformatica*, **11**(2): 217-237.

Cardoso J & A Escorcio 2007. Editing tools for ontology construction, in "*Semantic Web Services: Theory, Tools and Applications*", Idea Group.

Corcho O, Fernández-López M, Gómez-Pérez A, and O Vicente 2002. WebODE: an integrated workbench for ontology representation, reasoning and exchange. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management* (EKAW'02), pp. 138-153. Lecture Notes on Artificial Intelligence, 2473. Berlin: Springer-Verlag.

Da Silva G, Seixas R, and de Farias O 2005. Geographical Information Systems and Dynamic Modeling via Agent Based Systems. Proceedings of the ACM-Gis’05, 13th ACM International symposium on Advances in Geographical Information Systems, November 4-5, 2005, Bremen, Germany.

Duineveld A, Stoter R, Weiden M, Kenepa B and V Benjamins 2000. Wondertools - a comparative study of ontological engineering tools. *International Journal of Human-Computer Studies*, **52**(6):1111-1133.

ESRI 2008. *ArcGIS Data Models*. Web site at http://support.esri.com/index.cfm?fa=downloads.dataModels.gateway, last accessed 01/29/08.

Geertman S and J Stillwell (Eds) 2002. *Planning Support Systems in Practice*. Berlin: Springer.

Gimblett R 2002. *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes*. London: Oxford University Press.

GOL 2004. Georeference Online Ltd. Website at http://www.georeferenceonline.com, last accessed 01/27/08.

Gold C and S Cormack 1987. Spatially ordered networks and topographic reconstructions. *International Journal of Geographical Information Systems*, **1**:137-148.

Gómez-Pérez A, González Cabero R, and M Lama Peníin 2004. ODE SWS, a semantic web services development tool. In McIlraith S, Plexousakis D, and F van Harmelen (Eds.) The Semantic Web - ISWC 2004, *Proceedings of the Third International Semantic Web Conference*, Hiroshima, Japan, November 7-11, 2004. Lecture Notes in Computer Science , Vol. 3298 . Berlin: Springer.

Grenon P and B Smith 2003. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, **4**(1):69–103.

Groff E 2007a. 'Situating' Simulation to Model Human Spatio-Temporal Interactions: An Example Using Crime Events. *Transactions in GIS*, **11**(4):507-530.

Groff E 2007b. "Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery." *Journal of Quantitative Criminology*, **23**:75-103.

Groff E 2008. Adding the Temporal and Spatial Aspects of Routine Activities: A Further Test of Routine Activity Theory. *Security Journal*, **21**(1):95-116.

Huxhold W and A Levinsohn 1994. *Managing Geographic Information Systems*. Oxford University Press.

ISO/IEC JTC1/SC7 2005. WD 15909-2, Software and Systems Engineering, High-level Petri Nets. Montreal, Canada: ISO/IEC JTC1/SC7 Secretariat. Available online at http://wwwcs.uni-paderborn.de/cs/kindler/publications/copies/ISO-IEC15909-2-WD0.9.0.Ballot.pdf, last accessed 02/03/08.

Kemp K and A Vckovski 1998. Towards an ontology of fields. *Proceedings of the 3rd International Conference on GeoComputation*, University of Bristol, UK, 17-19 September 1998. Available online at http://www.geocomputation.org/1998/60/gc_60.htm, last accessed 5 May 2008.

Klien E and F Probst 2005. Requirements for Geospatial Ontology Engineering. *8th Conference on Geographic Information Science* (AGILE 2005). Estoril, Portugal.

Kuhn W 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, **15**:(7): 613-631.

Line M 2004. *Proposed Upper Ontology for the semiotics of Complex Systems*. Polymathix white paper, available online from http://www.polymathix.com/papers/socs-upper.html, last accessed 02/03/08.

Lüscher P, Burghardt D and R Weibel 2007. Ontology-driven Enrichment of Spatial Databases. *10th ICA Workshop on Generalisation and Multiple Representation*, 2-3 August 2007, Moscow.

Nogueras-Iso J, Zarazaga-Soria F, and P Muro-Medrano 2005. *Geographic Information Metadata for Spatial Data Infrastructures: resources, interoperability, and information retrieval*. Berlin: Springer.

North M, Collier N, and J Vos 2006. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation*, **16**:1–25.

Noy N and D McGuinness 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. Knowledge Systems Laboratory, March, 2001.

Obermeyer N and J Pinto 2007. *Managing Geographic Information Systems*. Guilford.

Peachavanish R & H Karimi 2007. Ontological Engineering for Interpreting Geospatial Queries, *Transactions in GIS*, **11**(1): 115-130.

PELLET 2008. Open source OWL DL reasoner web site at http://pellet.owldl.com/, last accessed 01/29/08.

Peuquet D, Smith B, and B Brogaard-Pederson 1999. Ontologies of Fields. *Varenius Project Specialist Meeting Report*, Santa Barbara, CA: National Center for Geographic Information and Analysis.

Peterson J 1981. *Petri Net Theory and the Modeling of Systems*. Prentice-Hall.

Petri C 1962. *Kommunikation mit Automaten*. PhD thesis, Bonn: Universität Bonn.

Protégé 2007. Protégé editor. http://protege.stanford.edu. Last accessed 01/26/08.

Qiu Q, Wu Q and M Pedram 2000. Dynamic Power Management of Complex Systems Using Generalized Stochastic Petri Nets. *Proceedings of the 37th ACM Design Automation Conference* (DAC2000), pp 352-56. Available online at http://www.sigda.org/Archives/ProceedingArchives/Dac/Dac2000/papers/2000/dac00/pdffiles/21_3.pdf, last accessed 02/03/08.

Raskin R 2004. *Guide to SWEET Ontologies*. Pasadena, CA: NASA / Jet Propulsion Lab.

Rational Rose 2003. *Using Rose*. Rational Software Corp. (now IBM).

Reitsma F and J Albrecht 2005. Implementing a New Data Model for Simulating Process. *International Journal for Geographic Information Science*, **19**(10):1073-90.

Selçuk B 2007. *Dynamic Performance of Hierarchical Planning Systems*. Eindhoven, NL: Technische Universiteit Eindhoven. Available online at http://alexandria.tue.nl/extra2/200612552.pdf, last accessed 02/03/08.

SemTalk 2008. Business process modeling and ontology editor. Potsdam, Germany: Semtation GmbH. URL www.semtalk.com, last accessed 01/29/08.

Sheng J, Sun X, Huag G, Jiao W, Sun Y and H Mei 2005. Towards a unified model for supporting mechanisms of dynamic component update. *ACM SIGSOFT Software Engineering Notes*, **30**(5): 80-89.

Souza D, Salgado A and P Tedesco 2006. Towards a Context Ontology for Data Integration. In Meersman R, Tari Z, and P Herrero (Eds) *On the Move to Meaningful Internet Systems*, LNCS 4278, pp 1576-1585, Berlin: Springer.

Theodorakis M, Analyti A, Constantopoulos P, and N Spyratos 1999. Contextualization as an Abstraction Mechanism for Conceptual Modeling. In Akoka J, Bouzeghoub M, Comyn-Wattiau I and E Metais (Eds), *Proceeding of the 18th International Conference on Conceptual Modeling (ER '99), Paris, France*, pp. 475-489. Lecture Notes in Computer Science 1728, Berlin: Springer.

Tomlinson R 2003. *Thinking About GIS: geographic information system planning for management*. Redlands, CA: ESRI Press.

Tsang H, Lay C and Y Leung 2004. *Object Oriented Technology: from diagram to code with visual paradigm for UML*. Princeton, NJ: McGraw-Hill.

W3C 2004. OWL-S: semantic markup for web services. W3C member submission 22 Nov. 2004. Online from: http://www.w3.org/Submission/OWL-S, last accessed 01/26/08.

Watson R 2005. *Data management: databases and organizations*. Wiley.

WebODE 2003. Ontological engineering platform, downloadable from http://webode.dia.fi.upm.es/WebODEWeb. Last accessed 01/26/08.

Williamson I, Rajabifard A, and M Feeney (Eds) 2003. *Developing Spatial Data Infrastructures: from concepts to reality*. London: Taylor & Francis.

Wootton C 2007. *Developing Quality Metadata: building innovative tools and workflow solutions*. Oxford, UK: Focal Press.

Yeung A and B Hall 2007. *Spatial Database Systems: design, implementation and project management*. Berlin: Springer.