

## **SIMPHO: AN ONTOLOGY FOR SIMULATION MODELING OF POPULATION HEALTH**

Anya Okhmatovskaia  
David L. Buckeridge  
Arash Shaban-Nejad  
Andrew Sutcliffe

McGill University  
1140 Pine Ave. West  
H3A 1A3 Montreal, QC, CANADA

Jacek A. Kopec

Arthritis Research Centre of Canada  
895 West 10th Avenue  
V5Z 1L7 Vancouver, BC, CANADA

Philippe Finès

Statistics Canada  
100 Promenade Tunney's Pasture  
K1A 0T6 Ottawa, ON, CANADA

Michael C. Wolfson

University of Ottawa  
451 Smyth Rd.  
K1H 8M5 Ottawa, ON, CANADA

### **ABSTRACT**

Simulation modeling of population health is being used increasingly for epidemiology research and public health policy-making. However, the impact of population health simulation models is inhibited by their complexity and the lack of established standards to describe these models. To address this issue, we are developing the Ontology for Simulation Modeling of Population Health (SimPHO) – a formal, explicit, computer-readable approach to describing population health simulation models. SimPHO builds on previous work to classify and formally represent knowledge about simulation models, and incorporates the semantics of the epidemiology and public health domains. SimPHO will allow model developers to make explicit their assumptions, to describe their models in a formal, consistent and interoperable manner, and to facilitate model reuse and integration. To illustrate the use of SimPHO, we describe one software application driven by this ontology, an automated visualization tool for generating interactive web-based diagrams of population health simulation models.

### **1 INTRODUCTION**

Computer simulation of population health is a powerful technique that has been used for over two decades by epidemiologists and public health researchers to study population-level health phenomena emerging from complex interactions of mechanisms at the levels of individuals, populations, and healthcare systems. With recent advances in computing technology and the increasing availability of electronic health data, simulation modeling continues to gain credibility and exert greater influence on public health policy. Modern applications of health simulation models include predicting transmission patterns of infectious diseases, estimating the economic burden of disease outbreaks, analyzing the performance of public health surveillance systems and evaluating the impacts of health interventions and policies.

There remain, however, significant issues that inhibit wider acceptance of agent-based models of population health among researchers and practitioners. Particularly, the inherent complexity of these models and the lack of standards for describing them present formidable barriers to communicating the details of the models between model developers on one side and other researchers and policy-makers on the other. Scientific publications presenting health simulation models vary in the accuracy and consistency

cy of documentation, and the level of detail is often insufficient to understand fully how the model works and what underlying assumptions were made in the process of its development. Low transparency of the models makes them almost impossible to reproduce and validate, and is also a barrier to their use in research and policy-making, to the extent that only a small proportion of the existing health simulation models are being used outside of the group that developed them.

A notable example of a large effort to address this problem is the Standardized Model Documents developed by Cancer Intervention and Surveillance Modeling Network (CISNET 2011). Unfortunately, the descriptions of cancer simulation models that are published on the CISNET website and comply with the proposed standard, still lack critical details, which would allow the user to understand or evaluate, not to mention replicate these models, which implies that this standard is insufficiently formal and/or complete.

To address the issues outlined above, we are developing an Ontology for Simulation Modeling of Population Health (SimPHO) – a software ontology to encode formally and explicitly population health simulation models, and a set of accompanying applications to create automatically documentation and diagrams intended to make these models more transparent and comparable. Our hope is that SimPHO will eventually serve as a formal standard for describing simulation models of population health, and that the software tools powered by it will be used by model developers, as well as model users in epidemiology research and public health practice.

## 2 BACKGROUND AND RELATED WORK

The problem of complexity of simulation models, the lack of interoperability and the need for standards to describe them is not new and has been widely acknowledged (Lorek and Sonnenschein 1999; Grimm et al. 2006; Silver et al. 2010). In this section we discuss previously suggested solutions to this problem and how they apply to simulation models of population health. Any reasonable book on modeling and simulation features some sort of classification, or a *taxonomy* of simulation models, as well as provides a *vocabulary* (glossary) explaining the meanings of important terms in natural language. A software *ontology* (to be distinguished from the philosophical meaning of this word) is a formal, explicit machine-readable specification of a domain of knowledge, which combines and builds upon aspects of a taxonomy and a vocabulary in a form of logical axioms. Ontologies arrange concepts in a taxonomic structure and formally define the properties and the relationships between these concepts.

Representing knowledge about simulation models constitutes a form of meta-modeling. In order to have any significant practical impact, simulation modeling requires accurate information about the entities, events, processes and their relationships within a real-world domain that is being simulated. Therefore, when introducing the means to formally and completely describe simulation models, we need to distinguish between what is known about the model itself, and about what that model represents.

### 2.1 Taxonomies and Ontologies for Modeling and Simulation

Modeling and simulation have been growing rapidly for several decades. Due to the intersection with numerous disciplines and application domains, simulation modeling today draws upon diverse approaches and methods, and consequently there is no single accepted taxonomy of models and modeling techniques.

A variety of classification criteria for simulation models exist. The models can be differentiated according to how they treat time (dynamic vs. static models), whether they incorporate noise or randomness (deterministic vs. stochastic models), and whether the changes in the state of the system are discrete or continuous (Law and Kelton 1999). Brennan, Chick, and Davies (2006) analyzed the models used for economic evaluation of health, and proposed to use other dimensions, such as the unit of analysis (individual-level vs. cohort/aggregate-level), whether interactions between simulated entities are modeled or not, how time is incorporated, and how the current state depends upon historical states (i.e. first-order vs. higher-order Markov models). Spielauer (2009) in his overview of microsimulation approaches suggests taking into account whether the model uses an open versus a closed population, and how the population was initialized (synthetic vs. cross-sectional starting population). He also distinguishes between discrete-

time and continuous-time models according to the mechanism of scheduling the events, and between explanatory and predictive models based on the model purpose.

Miller et al. (2004) presented a thorough review of the existing taxonomies for discrete event (DE) models and used the results of this review to develop an ontology for conceptualizing these models – the Discrete Event Modeling Ontology (DeMO) (Silver et al. 2010; Miller et al. 2004). DeMO presents a formal, consistent approach to distinguishing between the subclasses of DE models and defines the relations between each type of model and simulation modeling components and primitives (states, events, activities, transitions, etc.). DeMO developers have demonstrated how their ontology can be used to facilitate model development and discovery and to formalize the documentation of the model development process. While being a useful and well designed reference ontology, DeMO, however, is not ideally suited for describing the actual existing simulation models: many of them, particularly the models of population health, use a combination of modeling approaches, and cannot be formally classified using DeMO.

Another ontology – the ontology for agent-based simulation modeling (ABMS), was described by Christley, Xiang, and Madey (2004). In their work, the authors conceptualized not only the simulation models, but also the processes of model development, implementation and validation. The inclusion of model development is useful, since the details of model development and implementation are important for understanding a model and virtually indispensable for cross-model comparison. For example, the ontology for ABMS allows one to specify the inferred assumptions or the sources of model parameters.

Both ontologies mentioned above are domain-independent, and therefore can characterize simulation models in terms of structure using general simulation modeling terms. In many cases, however, describing the content of the model in the specific context of the simulated domain can facilitate understanding of the model, as well as model comparison and validation by domain experts. In our work to develop the SimPHO ontology, we have taken into account the specific applications of simulation modeling in population health and incorporated the relevant knowledge from this domain.

## **2.2 Simulation Modeling of Health**

While the impact of simulation modeling in medicine and health-care has not been as profound as in other domains such as tax and welfare policy, cosmology, high energy physics, and military and logistics applications, simulation models are used widely and they continue to gain popularity among epidemiologists and medical and public health professionals. A number of reviews have been published summarizing the use of simulation modeling in health care sector and identifying the associated challenges of both theoretical and applied nature (Brailsford 2007; Barjis 2010; Katsaliaki and Mustafee 2010). One recent survey outlined major application areas of simulation modeling in the health care sector: epidemiology, health promotion and disease prevention; health care systems operation; health care systems design; medical decision making; and extreme events planning (Mielczarek and Uziarko-Mydlikowska 2010). This survey also identified that most simulation research focused on evaluating the performance of health care systems and the effects of health policy.

In order for health simulation models to be used effectively in these application areas, the models must reflect a critical aspect of the health domain: being complex and multi-disciplinary, health domain necessitates a multi-level, multi-paradigm approach in modeling (Barjis 2010). For example, to accurately evaluate the health effects of certain policies at the population level, the model may need to incorporate a range of diverse factors and processes, such as the prevalence of risk factors and the observed trends in the upstream health determinants, the dynamics of disease progression and spread, human behavior in relation to health care systems, existing protocols for clinical care and reporting, socio-economic determinants, geographical and climatic factors, etc.

The inherent complexity of population health simulation models presents a challenge to their adoption by a heterogeneous user community with diverse background (Brailsford 2007; Barjis 2010). Combined with a lack of documentation standards, this complexity translates into poor transparency, which largely explains the low acceptance of simulation modeling in health domain, particularly given its high potential and relevance for this domain. The development of a standard ontology providing a conceptual frame-

work and software tools to formally, explicitly and consistently describe simulation models of population health is essential for promoting the wide use of these models. This ontology would have to account for the multidisciplinary nature of the field, as well as the diversity of the user community, relying on the established standards and terminologies, with which both model developers and potential users are familiar.

### **2.3 Ontologies in Health and Medicine**

The last decade has seen an incredible proliferation of biomedical ontologies (see Rubin, Shah, and Noy (2008) for review). While clear guidelines for ontology development exist, as well as the criteria for ontology evaluation, the quality of existing biomedical ontologies varies tremendously, and their reusability is often very limited. The institutionalization of biomedical ontologies and the establishment of OBO (Open Biomedical Ontologies) consortium were meant to coordinate ontology development in the field and to ensure some sort of quality standard.

The use of ontologies in health and medicine has mainly focused on the formalization and reorganization of medical terminologies and taxonomies. Some of these taxonomies, for example the International Classification of Diseases (WHO 2010), have long been established as industry standards, while the recently increased popularity of ontologies has spurred the efforts to encode these taxonomies in some formal logic broaden the range of their software applications. The issue with the majority of existing medical terminologies, despite their long history and extensive use, is that they were not initially designed as ontologies and often contain various logical and ontological errors and inconsistencies (Schulz et al. 2010), for example the common conflation of taxonomical (is-a) and mereological (part-of) relationships. Another common problem in large “ontologized” medical classification systems is unnecessary proliferation of terms and concept duplication. Their direct use in formal application ontologies, such as SimPHO, is therefore problematic. Alternatives to the well-known medical terminologies, which conform to the ontology standards, do exist, for example the Human Disease Ontology (Schröml et al. 2012). However their development is still in the early stages, and their acceptance by the medical community is minimal.

As stated in the previous subsection, it is critical for the health simulation modeling ontology to rely on standard terminologies in order to be practically useful. It is clear, however, that no existing medical ontology or standard terminology satisfies the criteria of widespread acceptance and high ontology quality standards at the same time. In our work in developing SimPHO we therefore have to combine internally consistent fragments of multiple existing medical ontologies and vocabularies.

## **3 SIMPHO OVERVIEW**

SimPHO stands for an Ontology for Simulation modeling of Population Health, and is a formal ontology, which defines concepts relevant and necessary for the unambiguous description of health simulation models at several conceptual levels. SimPHO is an application ontology, meaning that it was designed primarily to power a set of specific software tools (applications), as opposed to creating a general theory of the health simulation modeling domain, which would characterize a reference ontology (Menzel 2003).

We made an extensive use of expert knowledge when developing SimPHO. This work is part of the effort to leverage the use of simulation modeling technology for evidence-based decision-making at different levels of public health and health-care systems, which is led by the Simulation Technology for Applied Research (STAR) team. Our team is a Canadian network of health modelers and decision makers, based on a partnership among leading investigators from major academic centers, experts from Statistics Canada, and senior decision makers at the federal, provincial and health region levels.

SimPHO was implemented in OWL (Web Ontology Language) version 2 (W3C 2009) using Protégé-OWL editor (Stanford Center for Biomedical Informatics Research 2003).

### **3.1 Purpose and Intended Uses of SimPHO**

The overarching goal of developing SimPHO is to increase the transparency of health simulation models, which will facilitate their review, comparison and validation, and eventually improve the quality of these

models and their acceptance by the user community. It will also aid in the development of future simulation models of population health. To accomplish this goal, SimPHO is intended to provide a conceptual framework for describing health simulation models in a standard, consistent and machine-interpretable way. Individual simulation models, formally encoded in terms of SimPHO, will constitute what can be called a population health simulation modeling knowledge base, which would allow structured retrieval of information about simulation models and application of logical reasoning algorithms to this knowledge.

Following below is a list of sample queries, or questions one may be interested to ask about some specific simulation model(s), and that could be answered with the help of the SimPHO knowledge base:

What kind of population entities are the focus of the model?

What time period(s) are covered in the model?

Which socio-economic factors does the model account for, if any?

Which scale is used to represent/measure smoking in the model?

What are all the factors that affect incidence of heart disease in the model?

What is the functional form of the link between smoking and heart disease in the model?

In addition to structured retrieval of knowledge, formally representing simulation models in terms of SimPHO enables some more specific practical tasks, such as:

- automatically generate explicit model description for scientific publications and other media
- produce standard technical model documentation at varying level of detail, matching the needs of the specific user category (e.g. public health practitioners, epidemiologists, model developers)
- visualize various aspects of the model in a form of diagrams using existing graphical notations
- evaluate population health simulation models in terms of internal consistency, inclusion of relevant factors, adequacy of the utilized data sources, etc.
- enable replication and comparison of published simulation models and their results

To accomplish these tasks, we are developing a suite of specialized software applications, each of which will access the multifaceted knowledge encoded in SimPHO with the help of a standard application programming interface (API) and transform it according to their specific purpose. We describe one of these applications later in this paper.

While automatic code generation for simulation model composition and deployment has been a widely discussed area of ontology application, we believe that it is only feasible under a very restrictive set of conditions, e.g. for models with specific structure or using a single canonic modeling approach. The purpose of SimPHO, however, is to provide a flexible and general way of describing a wide range of diverse models. We therefore do not consider automatic code generation among its intended applications at present moment, although the structure and content of SimPHO in no way preclude adding this functionality.

### **3.2 SimPHO Domain Coverage and Scope**

As emphasized earlier, simulation modeling of population health is a highly multidisciplinary field. SimPHO, therefore, must encode a relatively large body of diverse knowledge. Following one of the key principles in knowledge engineering, we have attempted to reuse existing ontologies and other knowledge sources wherever it was appropriate. Table 1 enumerates knowledge domains incorporated in SimPHO to varying extent, along with the referenced sources.

In its current version SimPHO is primarily focused on representing discrete event simulations, since this class of models appears to be the most commonly used for modeling population health. More specifically, SimPHO currently focuses on microsimulation models of population health, which operate at the level of individuals (persons) and simulate the dynamics of health-related states and events for those individuals (including the models with interacting or non-interacting individuals). The depth is limited to representing state variables, specific mathematical equations, logical rules or mapping functions, but leaving out the programming language-specific details of software implementation.

In terms of the content of the simulation models, i.e. what is being modeled, SimPHO covers a wide range of diseases, their risk factors and outcomes, demographic characteristics of the population, health-

care associated costs and measures of disease burden. The models represented by SimPHO are epidemiological/population health models; models of diseases at physiological or biochemical level are outside the scope of the ontology. Following the rationale suggested by Mielczarek and Uzialko-Mydlowska (2010), we also exclude clinical simulations and training environments.

Table 1: SimPHO domain coverage and referenced knowledge sources

Domain	Knowledge source
modeling and simulation	<ul style="list-style-type: none"> <li>• DeMO (Silver et al. 2010)</li> <li>• Ontology for ABMS (Christley, Xiang, and Madey 2004)</li> </ul>
measurement and statistics	<ul style="list-style-type: none"> <li>• Ontology of Clinical Research (OCRe) (Tu et al. 2009)</li> <li>• Quantities, Units, Dimensions and Data Types (QUDT) (Hodgson and Keller 2011)</li> </ul>
health and medicine	<ul style="list-style-type: none"> <li>• Public Health Ontology (PHont) (developed internally, see further text)</li> <li>• Disease Ontology (Schrml et al. 2012)</li> <li>• International Classification of Diseases (ICD-10) (WHO 2010)</li> </ul>
time	<ul style="list-style-type: none"> <li>• OWL-Time ontology (W3C 2006)</li> </ul>
geography	<ul style="list-style-type: none"> <li>• Geography ontology (developed internally)</li> </ul>

### 3.3 Characterizing Health Simulation Models Using SimPHO

Based on the existing tradition of describing health simulation models in scientific literature, expert opinion and the proposed more formalized protocols (Grimm et al. 2006; CISNET 2011), we believe that a clear and complete description of a simulation model should consist of 3 major parts: the general high-level definition (what kind of model this is); the characterization of the content of the model in terms of the relevant domain of knowledge (what is being modeled); and the technical specification of modeling details (how it is modeled). A specific simulation model can be encoded as an instance of a “simulation model” concept (class), while model components, properties, specific mechanisms, etc. are also instantiated from other classes defined in SimPHO.

#### 3.3.1 General Description

The high-level description of the model in SimPHO relies on the commonly used classification dimensions, introduced in section 2.1. SimPHO defines a top-level class *SimulationModel* and a number of subclasses to represent different model categories. These subclasses are defined using logical axioms and class properties, which represent the relevant classification dimensions (e.g. *hasAnalyticUnit*) and logical axioms (Figure 1). In contrast to DeMO, the subclasses of *SimulationModel* in SimPHO are not necessarily disjoint due to a multi-axial nature of classification, so it is possible, for example, to encode a specific model as an instance of both *ContinuousTimeSimulationModel* and *MicrosimulationModel*. If needed, one can also use part of the DeMO taxonomy imported to SimPHO to categorize the model more specifically as, for example, a *DiscreteTimeMarkovChain*.

As part of high-level model description, SimPHO also allows one to specify the type of simulated population (open vs. closed) and the general method used for population initialization, as well as to identify the sub-components or modules of the simulation model in case the model is naturally decomposable into meaningful units. Finally, SimPHO provides a taxonomy of applications of health simulation models, which allows to characterize the model according to its purpose.

#### 3.3.2 Content of the Model in Health Domain Terms

A simulation model is a representation of a real world system, and the entities, relationships and processes within the simulation model map onto corresponding objects, relationships and processes in the modeled domain. The domain-specific knowledge does not only help the user of the model make sense of the sim-

ulation procedures and results, but also makes it possible to validate the model. Silver et al. (2010) have even demonstrated how the knowledge resident in the domain ontologies can possibly be integrated with the knowledge in the modeling ontology (DeMO) to construct executable simulation models.

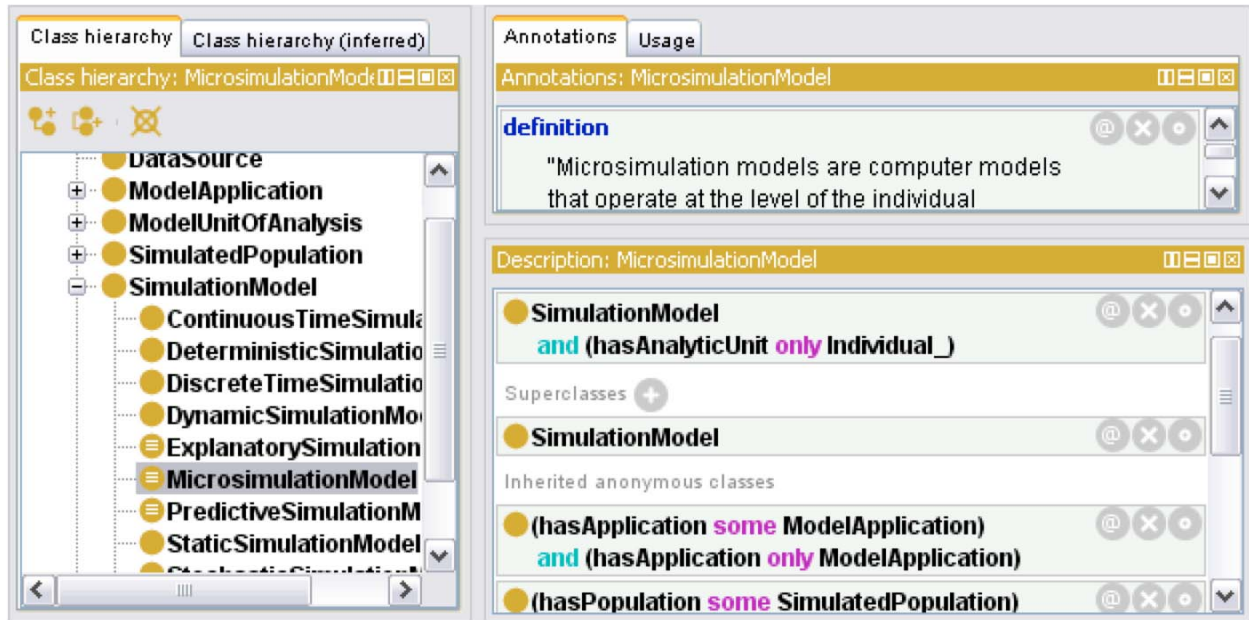


Figure 1: Definition of simulation model types in SimPHO (a screen-shot from the Protégé editor)

Given the current focus of SimPHO on the multi-level health models used for evaluation of health-care systems and public health policies, we have selected and structured the domain-specific knowledge to emphasize the essential concepts and relationships relevant for this application area. More specifically, the ontology should define individual and population-level characteristics indicative of and affecting the health status, health issues and events that are important from the point of view of health-care systems and policies, identify the known causal health effects and describe any other relevant factors, which might be relevant to simulation modeling of health (e.g. environmental, geospatial, institutional, etc.) – the knowledge, which is typically associated with the domain of public health.

There is no readily available OWL ontology of public health, however an Australian team of researchers has recently proposed a taxonomy of public health specific to their country (Jorm, Gruszyn, and Churches 2009). We have adapted and generalized their taxonomy to build our own Public Health ontology (PHont), which is currently in the stage of active development by our research group. PHont provides a detailed classification within each of the following broad categories: determinants of health and risk factors, health issues, public health methods and interventions, public health resources, and public health settings. It also identifies the properties of core concepts and the relationships between them. In particular, known health effects of risk factors specific to certain population groups are incorporated in PHont along with the knowledge about modifiability of these risk factors. Note that this knowledge is independent from how specific causal relationships between risk factors and health issues are modeled within a given simulation model, and serves as a reference allowing the development of automatic tools for model assessment and validation. For example, smoking is known to be a risk factor for acute myocardial infarction (AMI), but for various reasons it may or may not be included in a given simulation model of AMI, and if included, the relationship can be modeled as direct or indirect. These specifics of modeling cannot be meaningfully evaluated as advantages or limitations without referring to the domain knowledge. PHont is a reference ontology, which is imported into SimPHO to supply this essential domain knowledge.

In addition, SimPHO uses several other sources of domain knowledge (see Table 1), to conceptualize a variety of diseases, their progression, complications and outcomes, treatment protocols, characterize

human health-care seeking behavior, and the logistics of healthcare encounters. These concepts add domain-specific meaning to simulated events, processes, actions and the states of simulated entities.

### 3.3.3 Technical Details of the Modeling

This part of the simulation model description provides the computational details about how the model works, and, to the extent that it may be relevant for cross-model comparison and validation – how it was developed. Such description allows one to unambiguously encode the internal functioning of a simulation model, to completely specify relationships among states and events in a simulation (so that a model can be re-implemented according to this specification), and to drive reasoning over encoded simulation models allowing for consistency checking and conversion to standard formalisms and graphical notations. To achieve these goals, SimPHO formalizes the following aspects of simulation models:

- data types and measurement scales of state variables
- procedures for initializing state variables
- model parameters, methods and data sources used for their derivation
- causal relationships between states and events, including some specific cases of causality (direct vs. indirect effects, effect modification, conditioning, enabling vs. triggering, etc.)
- temporal dynamics of the simulation model
- specific mathematical and statistical functions used to compute event probabilities, the values of state variables and the timing of events.

There is a relatively small number of modeling primitives, which SimPHO mainly borrows from existing ontologies (e.g. DeMO, the ontology for agent-based modeling and simulation). These concepts, nevertheless, provide a framework sufficient for describing the low-level details of the simulation mechanisms of the models in order to accomplish the goals outlined above. In addition, SimPHO uses the concepts from the supporting ontologies on statistics, units of measurement, and temporal relations.

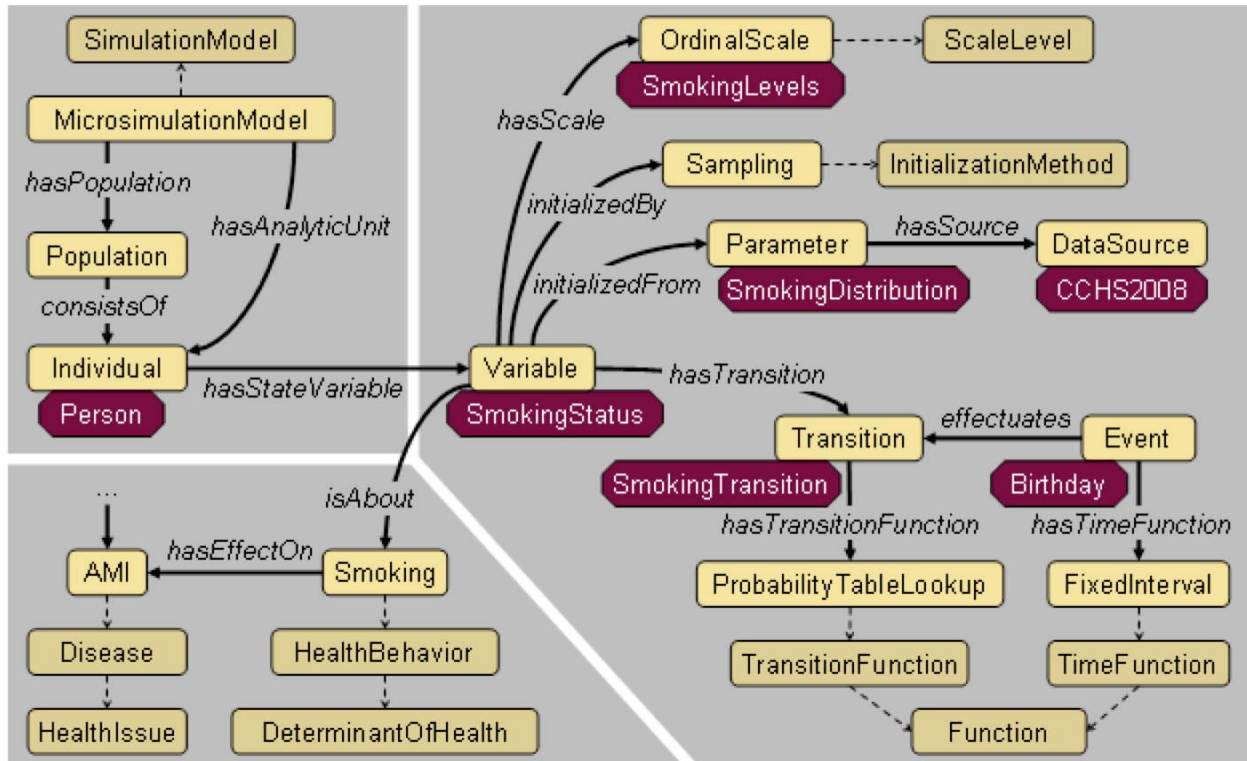


Figure 2: Example of simulation modeling concepts and their relationships in SimPHO



Figure 2 schematically depicts a small fragment of SimPHO illustrating how some of its concepts and their relationships can be used to encode the details of modeling smoking in a particular microsimulation model of population health. Rounded rectangles represent classes in SimPHO, octagons – instances specific to this simulation model, dashed lines denote taxonomical “is-a” relationship, while solid bold lines denote associative relationships in a form of class properties. In this simplified example, *SmokingStatus* is one of the state variables (instance of a *Variable* class) of a simulated individual *Person*. *SmokingStatus* is a categorical variable, which can be assigned one of ordered *SmokingLevels*. The initial values of *SmokingStatus* are sampled from a distribution of smoking levels (*SmokingDistribution*), which is a model *Parameter* that was derived from the Canadian Community Health Survey (CCHS) data. The graph also shows how evolution of *SmokingStatus* is modeled: the new value of this variable is evaluated at *Birthday* (instance of an *Event*), and is determined using a transition probability table; time interval before next *Birthday* event is fixed. It is also possible to specify the data source and stratification dimensions for the smoking transition probabilities (not shown on the graph). Figure 2 also shows some concepts used for general model description (top left section of the graph) and specification of model structure (bottom left), as explained in previous sections.

## 4 SIMPHO PROTOTYPE APPLICATION: OPHIUCHUS

As described in section 3.1, SimPHO is designed as an application ontology, which is created in parallel with a set of accompanying software tools intended to facilitate simulation model development, validation, comprehension and reuse. In this work we present a first prototype software application of SimPHO called Ophiuchus – a domain-specific visualization tool for generating interactive conceptual 2D diagrams of population health simulation models. Ophiuchus was implemented as a web-based application, relying on open-source tools and libraries, in order to make it readily available to a large community of users including epidemiologists, statisticians, public health researchers and practitioners.

### 4.1 Diagrammatic Representation of Simulation Models

The simulation modeling community has proposed standard protocols for describing simulation models (Grimm et al. 2006; CISNET 2011). These protocols, however, are insufficient for the purpose of model reproducibility or formal validation due to the flexibility and ambiguity of textual representation (Onggo 2011). Compared to text, formal diagrammatic representation supports more intuitive analysis and explicitly preserves the topological relations among the components of the reality that it represents, which is critical in the case of simulation models (Larkin and Simon 1987).

Most of the existing formalisms proposed for diagrammatic representation of simulation models (Onggo 2011) are specific to a certain class of models and/or represent only a particular aspect or view of the model. It is often not feasible to show all possible relationships within a model on a single diagram because of the inherent incompatibility of some views. Another well-studied problem with diagrammatic representation is graph size. Contemporary models of population health are often large and complex, and even when presented in a consistent and clear fashion, can be overwhelming. The approaches used in graph visualization to address this problem (Herman, Melancon, and Marshall 2000) require interactivity, so that only part of the information is displayed at a time, and details are provided on-demand.

### 4.2 Ophiuchus System

We used the existing body of knowledge on graph visualization when designing Ophiuchus to address the problems outlined above. Ophiuchus takes advantage of the comprehensive information about a simulation model formally encoded using SimPHO. At any given time, Ophiuchus retrieves and visualizes only a subset of encoded entities and relationships (e.g. causality, data dependencies, event sequencing, etc.), while providing additional information in an interactive manner. The retrieved set of relationships is specific to the type and application of the diagram being generated. For example, epidemiologists might be primarily interested in visualizing causal relationships between state variables using causal diagrams

(Greenland, Pearl, and Robins 1999). To address the problem of graph size and complexity, we employed known techniques to facilitate interactive incremental graph exploration, such as semantic zoom (more detailed information about the element in focus) and “focus + context” (persistent high-level contextual view of the graph with a reference to the element in focus) (Herman, Melancon, and Marshall 2000).

The system architecture and functionality of Ophiuchus are described in detail in Sutcliffe et al. (2012). At the back-end, the semantic relationships extracted from the ontology are translated to create a graph definition template, which is then used to perform automatic graph layout and generate 2D images by the Open Graph Drawing Framework (OGDF) C++ libraries (OGDF 2012). The front-end of the Ophiuchus is a web page that allows for interactive exploration of the generated model diagram: with the help of simple mouse actions, the users can retrieve additional information about the nodes and arcs of the graph, zoom in and out of node clusters (nodes grouped according to specific semantic criteria stored in SimPHO), hide and reveal parts of the graph, and change focus.

We have performed a preliminary informal assessment of the Ophiuchus system with a mixed group of twenty users. The user group included researchers and professionals working in the fields of epidemiology, public health, statistics and simulation modeling. Despite a few technical limitations of the system that were revealed in this evaluation study, we received positive feedback from the subjects, particularly regarding the interactive nature of the diagrams and the flexibility in navigating potentially large and complex graphs. Their overall enthusiasm about the application suggested that Ophiuchus could become quite helpful for the intended user-group by addressing the existing lack of specialized tools to describe and explain simulation models of population health in scientific and professional communication.

## 5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a first domain-specific application ontology for simulation modeling – SimPHO, which addresses the existing lack of standards for describing simulation models of population health. SimPHO provides a conceptual framework for clear, unambiguous, multifaceted representation of these models and serves as a foundation for a set of software tools intended to facilitate model comprehension, validation and reuse. Future work in the development of SimPHO will focus in four major areas. First, the scope of SimPHO will be extended to incorporate a wider range of simulation model types and applications. The structure of the ontology will also likely evolve to reflect any changes in the theory of modeling and simulation and other relevant knowledge domains SimPHO relies on. The second direction of future work is building the simulation modeling knowledge-base by formally encoding some of the existing health simulation models in terms of SimPHO. While representing all existing models is certainly beyond our aspirations, we will provide representative examples of model encoding to facilitate the adoption of SimPHO in the simulation modeling community. Experience encoding the models may also help refine structure of SimPHO. We have already started this work by encoding two distinct simulation models developed by the researchers from our group independently and using different modeling approaches. Thirdly, our future efforts will focus on further application development, as proposed in section 3.1. And finally, the ontology and accompanying applications will have to be formally and rigorously evaluated.

## REFERENCES

- Barjis, Joseph. 2010. “Healthcare Simulation and its Potential Areas and Future Trends.” *Simulation* 86 (8-9): 459-462.
- Brailsford, Sally C. 2007. “Advances and challenges in healthcare simulation modeling: tutorial.” In *Proceedings of the 2007 Winter Simulation Conference*, 1436-1448.
- Brennan, Alan, Stephen E Chick, and Ruth Davies. 2006. “A taxonomy of model structures for economic evaluation of health technologies.” *Health economics* 15 (12) (December): 1295-310.
- CISNET. 2011. Standardized Model Documents. <http://cisnet.cancer.gov/profiles/>.
- Christley, Scott, Xiaorong Xiang, and Greg Madey. 2004. “An ontology for agent-based modeling and simulation.” In *Proceedings of the Agent 2004 Conference*.

- Greenland, Sander, Judea Pearl, and James M. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10: 37-48.
- Grimm, Volker, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K. Heinz, and Geir Huse. 2006. "A standard protocol for describing individual-based and agent-based models." *Ecological Modelling* 198 (1-2): 115-126.
- Herman, I., G. Melancon, and M.S. Marshall. 2000. "Graph visualization and navigation in information visualization: A survey." *IEEE Transactions on Visualization and Computer Graphics* 6 (1): 24-43.
- Hodgson, Ralph, and Paul J. Keller. 2011. QUDT - Quantities, Units, Dimensions and Data Types in OWL and XML. <http://qudt.org/>.
- Jorm, Louisa, Su Gruszin, and Tim Churches. 2009. "A multidimensional classification of public health activity in Australia." *Australia and New Zealand health policy* 6 (1) (January): 9.
- Katsaliaki, K, and N Mustafee. 2010. "Applications of simulation within the healthcare context." *Journal of the Operational Research Society* 62 (8) (October 13): 1431-1451.
- Larkin, J, and H Simon. 1987. "Why a Diagram is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11 (1) (January): 65-100.
- Law, Averill, and W. David Kelton. 1999. *Simulation Modeling and Analysis (Industrial Engineering and Management Science Series)*. McGraw-Hill Science/Engineering/Math.
- Lorek, H., M. Sonnenschein. 1999. "Modeling and Simulation Software to support individual-oriented ecological modelling." *Ecological Modeling* 115: 199-216.
- Menzel, Christopher. 2003. "Reference Ontologies — Application Ontologies: Either/Or or Both/And?" In *Proceedings of the KI2003 workshop on Reference Ontologies and Application Ontologies*, 1-10.
- Mielczarek, B., and J. Uziarko-Mydlukowska. 2010. "Application of computer simulation modeling in the health care sector: a survey." *Simulation* 88 (2) (November 15): 197-216.
- Miller, John A., Gregory T. Baramidze, Amit P. Sheth, and P. A. Fishwick. 2004. "Investigating ontologies for Simulation modeling." In *Proceedings of the 37th Annual Simulation Symposium*, 55-71.
- OGDF. 2012. OGDF - Open Graph Drawing Framework. <http://www.ogdf.net>.
- Onggo, Stephan. 2011. "Methods for Conceptual Model Representation." In *Conceptual Modeling for Discrete-Event Simulation*, ed. S. Robinson, R. J. Brooks, K. Kotiadis, and D. van der Zee, 337-354.
- Rubin, Daniel L, Nigam H Shah, and Natalya F Noy. 2008. "Biomedical ontologies: a functional perspective." *Briefings in bioinformatics* 9 (1) (January): 75-90.
- Schriml, Lynn Marie, Cesar Arze, Suvana Nadendla, Yu-Wei Wayne Chang, Mark Mazaitis, Victor Felix, Gang Feng, and Warren Alden Kibbe. 2012. "Disease Ontology: a backbone for disease semantic integration." *Nucleic acids research* 40 (Database issue) (January): D940-6.
- Schulz, Stefan, Daniel Schober, Ilinca Tudose, and Holger Stenzhorn. 2010. "The Pitfalls of Thesaurus Ontologization - the Case of the NCI Thesaurus." *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2010* (January): 727-31.
- Silver, G. a., J. a. Miller, M. Hybinette, G. Baramidze, and W. S. York. 2010. "DeMO: An Ontology for Discrete-event Modeling and Simulation." *Simulation* 87 (9) (December 8): 747-773.
- Spielauer, Martin. 2009. *Microsimulation Approaches*. Statistics Canada working paper.
- Stanford Center for Biomedical Informatics Research. 2003. Protégé. <http://protege.stanford.edu>.
- Sutcliffe, A., A. Okhmatovskaia, A. Shaban-Nejad, D. Buckeridge. 2012. "Ophiucus: RDF-Based Visualization Tool for Health Simulation Models." In *Proceedings of the XXIV Conference of the European Federation of Medical Informatics*.
- Tu, Samson W, Simona Carini, Alan Rector, Peter Maccallum, Igor Toujilov, Steve Harris, and Ida Sim. 2009. "OCRe: An Ontology of Clinical Research." In *11th International Protégé Conference*.
- W3C. 2006. Time ontology in owl. <http://www.w3.org/TR/owl-time/>.
- W3C. 2009. OWL 2 Web Ontology Language Overview. <http://www.w3.org/TR/owl2-overview/>.
- WHO. 2010. International Classification of Diseases 10th revision. World Health Organization. <http://www.who.int/classifications/icd/en/>.

## **AUTHOR BIOGRAPHIES**

**ANYA OKHMATOVSKAIA** has a Ph.D. in Psychology from Moscow University, and a M.S. in Computer Science from the University of Southern California. Between 2004 and 2006 she was working as a Research Assistant at the USC Institute for Creative Technologies in the area of computational modeling of emotion. Since 2006 Anya Okhmatovskaia is a research associate at the McGill Clinical and Health Informatics Research group. Her main interests lie in knowledge modeling and agent-based simulation of population health. Her email address is [anna.okhmatovskaia@mcgill.ca](mailto:anna.okhmatovskaia@mcgill.ca).

**DAVID L. BUCKERIDGE** is an Associate Professor of Epidemiology and Biostatistics at McGill University where he holds a Canada Research Chair in Public Health. He is also a Medical Consultant to the Montreal Public Health Department and the Quebec Public Health Institute. His research focuses on the informatics of public health surveillance and disease control. He has a M.D. from Queen's University, a M.Sc. in Epidemiology from the University of Toronto, and a Ph.D. in Biomedical informatics from Stanford University. Dr Buckeridge is a Fellow of the Royal College of Physicians and Surgeons of Canada with training in Public Health and Preventive Medicine. His email address is [david.buckeridge@mcgill.ca](mailto:david.buckeridge@mcgill.ca).

**ARASH SHABAN-NEJAD** is a Postdoctoral Fellow at the Clinical & Health Informatics Centre at McGill University, where he conducts research on the knowledge representation and modeling in healthcare. He received his M.S. and Ph.D. in Computer Science from Concordia. His primary research interest is knowledge representation and semantic web, particularly ontologies and knowledge bases, description logics, category theory, agents, reasoning and inference, with special emphasis on applications in health and biomedical domains. His email address is [arash.shaban-nejad@mcgill.ca](mailto:arash.shaban-nejad@mcgill.ca).

**ANDREW SUTCLIFFE** is an undergraduate student at McGill University working towards Joint Bachelors in Mathematics and Computer Science. His interest are mainly in Human-Machine Interaction and Interaction Design. Since 2011 Andrew Sutcliffe has been working with the McGill Clinical & Health Informatics Research group, where he is developing ontology-driven software applications. His email address is [andrew.sutcliffe@mail.mcgill.ca](mailto:andrew.sutcliffe@mail.mcgill.ca).

**PHILIPPE FINES** has a Ph. D. in Statistics from Université de Sherbrooke, Canada. He is a senior analyst at Statistics Canada. He has a vast experience on statistical modeling and microsimulation in health. His current work involves development of microsimulation models of osteoarthritis and cancer, models of survival analysis and methods on hospitalization data. His email address is [philippe.fines@a.statcan.gc.ca](mailto:philippe.fines@a.statcan.gc.ca).

**JACEK A. KOPEC** is a professor in the School of Population and Public Health, UBC. He is also a Senior Scientist at the Arthritis Research Centre of Canada and is affiliated with the UBC Department of Medicine, the Centre for Hip Health and Mobility, and the National Surgical Adjuvant Breast and Bowel Project in the USA. Dr. Kopec earned his MD degree from the Pomeranian Medical University in Szczecin, Poland, and his PhD from the Department of Epidemiology and Biostatistics, McGill University. His interests include musculoskeletal epidemiology, quality of life measurement, and applications of simulation models in population health. His email address is [jkopec@arthritisresearch.ca](mailto:jkopec@arthritisresearch.ca).

**MICHAEL C. WOLFSON** is the Canada Research Chair in Population Health Modelling/Populomics at University of Ottawa. He is a former Assistant Chief Statistician, Analysis and Development, at Statistics Canada, and was a Fellow of the Canadian Institute of Advanced Research from 1988 to 2003. His recent research interests include analysis of the determinants of health, design of health information systems, income distribution, tax/transfer and pension policy analysis, and microsimulation approaches to socio-economic accounting and to evolutionary economic theory. Dr. Wolfson received his B.Sc. in computer science and economics from the University of Toronto, and his Ph.D. in economics from Cambridge University in 1977. His email address is [michael.wolfson@uottawa.ca](mailto:michael.wolfson@uottawa.ca).