Towards an Ontology for an Epidemiological Monitoring System

Gaoussou Camara
LANI, Université Gaston Berger, B.P. 234, Saint-Louis du Sénégal
gaoussoucamara@gmail.com

Sylvie Despres
LIM&BIO, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny
sylvie.despres@univ-paris13.fr

Rim Djedidi
LIM&BIO, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny
rim.jedidi@univ-paris13.fr

Moussa Lo
LANI, Université Gaston Berger, B.P. 234, Saint-Louis du Sénégal
moussa.lo@ugb.edu.sn

ABSTRACT

Epidemiological monitoring systems are used to control the evolution of disease spreading and to suggest action plans to prevent identified risks. In this domain, risk prediction is based on quantitative approaches that are hardly usable when data collection is not possible. In this paper, a qualitative approach based on an epidemiological monitoring ontology is proposed. We describe the design of this ontology and show how it fits into classical monitoring systems and helps overcoming limits related to quantitative approaches.

Keywords

Epidemiological Monitoring, Disease Control, Risk Analysis and Early Prediction, Simulation, Ontological Modeling.

INTRODUCTION

Monitoring systems are used to control epidemiological phenomena. They aim to predict event impacts constituting risk factors for disease spread, to control their evolution, and to suggest action plans to prevent identified risks. As noted in (Buton, 2006), monitoring is control to act in advance. Monitoring involves, in its implementation process, several organizations and actors coming from different disciplines.

Quantitative risk prediction is driven by simulation using numerical models. These models built from descriptive surveys, help explaining the dynamics of disease spread and validating assumptions. However, they are hardly useable for prediction and decision-support purposes in monitoring context. Indeed, simulation models are either designed for very limited target, or require inputs that are difficult to acquire in real-time simulations. Further, the diversity of formalisms used to represent these models (regular differential equations, agent-based models, etc.) restricts model composition and interoperability that are needed to answer to complex queries. Furthermore, to communicate, collaborate and share domain knowledge, the different organizations and actors involved need a common vocabulary. To settle these issues, an ontology-based qualitative approach is proposed for epidemiological monitoring system. An ontology is a formal and explicit specification of a shared conceptualization (Studer, Richard Benjamins and Fensel, 1998). Using an ontology for epidemiological monitoring allows: (i) qualitative simulations that can replace numerical simulations when data acquisition is difficult (Forbus, 1997), (ii) specification of knowledge related to numerical simulation models to facilitate model composition and interoperability, and provides: (iii) support for explaining observation and implementing action plans, and (iv) a common vocabulary to share knowledge and improve communication between organizations and actors involved in the disease monitoring process.

In this paper, we describe the design of an ontology for epidemiological monitoring and show how it fits into classical monitoring systems and helps overcoming limits related to quantitative approaches.The paper is organized as follows. In section 2, we present the epidemiological monitoring and the limits related to its implementation. In section 3 we detail the proposed ontology-based approach then, we illustrate in section 4, our contribution in the context of schistosomiasis monitoring in Senegal. In last section, we conclude and present future work.

EPIDEMIOLOGICAL MONITORING

Monitoring is an activity that consists in continuous surveillance of a phenomenon in order to anticipate and control its evolution. When it comes to epidemiological monitoring, the monitored phenomenon is the spread of an epidemic into a population of a specified area. Disease spreading is characterized by the evolution of the number of infected persons among a population. The risk of being contaminated is drastically increased by the presence of risk factors which influence the disease spread. These factors may be related to environmental, human or behavioral characteristics, among others. Thus, epidemiological monitoring consists in identifying and controlling these risk factors. When an event related to a risk factor is detected, all data necessary to analyze its impact upon the disease spread is collected. The analysis consists in handling numerical simulations in order to predict precisely the risks related to the occurrence of an event.

The analysis method used in monitoring systems depends on the nature of the studied phenomenon. Disease spread is considered as a system made of several entities (host, vector, pathogen agent, risk factors, etc.) whose interactions give rise to emergence of events (new disease cases, water infestation, etc.) at different levels of spatial and temporal scales. Epidemiological phenomena, owing to their evolution and their emergence, are considered as complex systems. To predict their evolution and dynamics, complex systems need to be analyzed through systemic modeling (Le Moigne, 1990). This allows giving a simplified view of the phenomenon (simulation model) in order to reproduce its behavior within a numerical simulation. Such simplification is attained through filtering (dimension lowering) adapted to the question that needs to be answered. Several systemic modeling approaches are interesting for studying complex systems, such as mathematical approaches and computer science approaches based on multi-agent systems (Amouroux, Taillandier and Drogoul, 2010; Drogoul, Treuil and Zucker, 2008). However, simulation requires the availability of appropriate simulation models and data in order to answer questions about the disease, such as “what is the possible evolution of the disease in coming weeks, since we have detected an event related to a spread risk factor?”. When the analysis of a detected event reveals a risk related to a disease, an alert is launched and decisions are made in anticipation.

Limits in implementing the epidemiological monitoring process

Data collection

Data collection is an essential phase for the analysis of the detected events by a monitoring system. Setting up a real-time automatic data collection system – to guarantee fast and accurate risk simulation and assessment – requires technical means that are not always available. A delay on the availability of data postpones simulation launch, and this may lead to consequences like a detected event appearing before being predicted. In some cases, collected data is not complete, leading to distorted quantitative analysis or misinterpretations of observations.

Simulation Handling

Simulation models are often reductionist. They model only a specific process for a specific target. Thus, they do not integrate all domain knowledge. To answer complex questions, simulation handing can require composition of several models. Nevertheless, available simulation models – that can answer to these complex questions – may be defined using different formalisms (regular differential equations, agent-based models, etc.) and thus, be hardly interoperable and composable. Further, two models can be expressed differently and yet have the same target. This leads to problems especially related to their selection, comparison and composition.

Communication and knowledge sharing

Epidemiological monitoring involves several organizations (health control and prevention cells, pharmaceutical industries, etc.) all working towards the same target, but often at different levels of action (environmental-level, population-level, drug-manufacturing, etc.). The monitoring system should enable knowledge sharing between all these structures. Furthermore, actors participating in risk analysis and decision-making may have heterogeneous profiles (epidemiologist, biostatistician, pathologist, meteorologist, doctor, entomologist, parasitologist, public health actor, political actor, etc.) and use different vocabularies for the same domain concepts. Thus, defining a common vocabulary is required to facilitate communication, interaction and collaborative work.

TOWARDS AN ONTOLOGY-BASED EPIDEMIOLOGICAL MONITORING SYSTEM

The purpose of this work is to propose an alternative and complementary approach to monitoring systems that handles monitoring process in different possible situations:

- Overcome constraints related to numerical data acquisition to perform simulation when early prediction of potential (or possible) risk is needed.
- Identify the source or cause of an observation when needed.
- Compose numerical simulation models when needed, and ensure their interoperability.
- Allow knowledge sharing when several organizations are involved in the process.
- Support communication and collaboration between different actors.

To address these challenging situations, the proposed approach is based on building an ontology for the epidemiological monitoring domain. Such an ontology provides:

- Qualitative specification of involved processes modeling the behavior of the phenomenon without numerical parameters. This allows reproducing disease spread process from the abstract description of its internal processes, and defining the different inherent states of the entities involved in the spread process. Thus, a qualitative simulation using ontological models will allow predicting states or occurrences of possible events related to a disease without numerical data.
- Common vocabulary of the domain to facilitate communication between actors.
- Explicit description of domain knowledge, including risk factors and existing simulation models. This description will allow knowledge sharing between involved organizations and ensure interoperability between simulation models.
- Formal representation enabling knowledge-based qualitative simulations that allows performing early predictions, guiding interpretation of observed phenomena or numerical simulation results, supporting decisions about action plans and driving their implementation.

Knowledge to model for building an epidemiological monitoring ontology

Epidemiological monitoring includes knowledge related to static aspects that can be modeled by classical domain ontology building approaches, and knowledge related to dynamic aspects corresponding to the different processes involved. Two ontological modeling approaches are thus used: domain knowledge modeling (knowledge on static aspects) and process knowledge modeling (knowledge on dynamic aspects).

Ontological distinction between domain knowledge and process knowledge

In philosophy and linguistics, there is a fundamental distinction between world static entities (known as continuants or endurants) and world dynamic entities (known as occurrents or perdurants). Perdurant concepts correspond to entities occurring over time with different phases, such as processes, events and states; endurant concepts correspond to entities without temporal part, such as objects (Lewis, 1986). They have been used in foundational ontologies, such as BFO (Grenon and Smith, 2004) and DOLCE[1], to guide the built of ontologies including static and dynamic domain aspects.

[1] http://www.loa.istc.cnr.it/ontologies/OWN/OWN.owl

Domain knowledge modeling

Ontological modeling of domain knowledge focuses on describing static entities and their relationships. It allows specifying knowledge about static entities in the domain regardless of their existence in time. It provides a common vocabulary of the domain to facilitate communication, and an explicit resources description (such as numerical simulation models) to support their sharing and ensure interoperability.

Process knowledge modeling

(Le Moigne, 1990) defines a process by its occurrence and its outcome. Numerical modeling approaches describe physical characteristics of processes while qualitative approaches, such as ontology-based approach, consider processes as concepts (abstraction). An ontological process model specifies classes of processes, relationships between processes, relationships between processes and objects, occurrence condition processes, process occurrence effects on states of other processes and objects, etc. The purpose of ontological modeling of processes is to allow reproducing possible behaviors of a system from the abstract description of its internal processes and its different possible states. Thus, reasoning on process ontology will allow predicting possible states or occurrences of a process, or explaining the cause of a process or a state occurrence.

Contents and organization of an epidemiological monitoring ontology

The concepts *process*, *event*, *state* and *object* and their abstract relations, such as *participation*, are high-level concepts and relations already modeled in foundational ontologies, such as DOLCE. We use these definitions to model static and dynamic aspects of epidemiological monitoring domain. The proposed ontology is structured into three layers (cf. figure 1): the foundational ontology layer, the core ontology layer and the specific ontology layer. The specific ontology layer models epidemiological monitoring specific characteristics. It is organized into two parts: static aspects corresponding to domain knowledge and dynamic aspects corresponding to knowledge about processes. The relationships between these two parts are also modeled. The core ontology layer contains common concepts of static and dynamic aspects according to ontology building methodology.

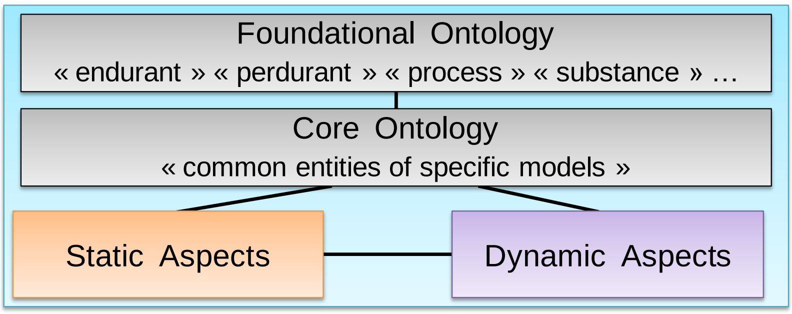


Figure 1. Ontology Structure

Architecture of an ontology-based epidemiological monitoring system

The architecture of the proposed monitoring system is based on the described epidemiological monitoring ontology (cf. *supra*). It includes a base of simulation models that are used by the monitoring support module and the numerical simulator (cf. figure 2).

Monitoring allows detecting the events related to disease risk factors. The monitoring support module carries on requests launched for analysis. If simulation models and data are not straight available, a qualitative simulation is performed based on ontological reasoning. Otherwise, the requests are processed by numerical simulations. In this case, the monitoring support module supervises the implementation of simulations by guiding the selection and composition of numerical simulation models. If a disease spreading risk is identified at the end of the monitoring process, an alert is triggered. Then, decisions are generated and evaluated by a quantitative or qualitative approach to prevent identified risks.

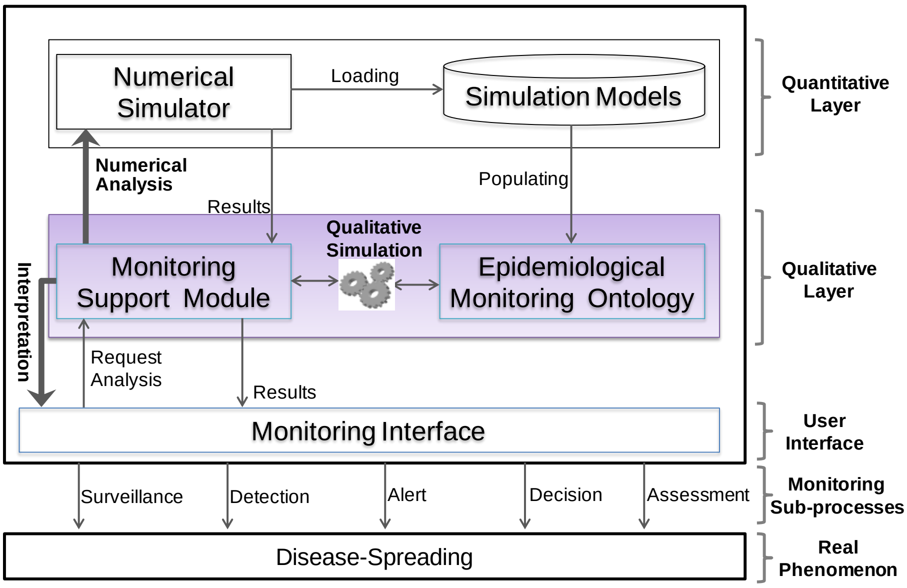


Figure 2. Ontology-based Epidemiological Monitoring System Architecture

APPLICATION FRAMEWORK: MONITORING OF SCHISTOSOMIASIS IN SENEGAL

In this section, we will show the contribution of modeling an ontology for epidemiological monitoring. Three use cases are defined for schistosomiasis monitoring in Senegal: early risk prediction, composition of simulation models, and support for explaining observations and implementing action plans.

Schistosomiasis in Senegal

Schistosomiasis (or bilharzia) is a parasitic disease constituting an important public health problem. Its contamination mode is based on indirect transmission. The infected person contaminates water by urine or feces containing parasite eggs. These eggs will hatch in fresh water if suitable conditions (temperature at 25 to 30 °C, sunshine, neutral pH) are met and then, larvae are released. These larvae are housed in gastropod mollusks and evolve to the stage of cercaria before being released into the water. These mollusks tend to prefer warm waters (22 to 28 °C), shaded, stagnant or moderate current and abundant vegetation (rivers, backwaters, ponds, natural and artificial lakes, irrigation networks) and they usually live 20 to 30 cm deep on plant stems, dead leaves or in mud bottom. With water contact, a healthy person is infected with skin cercariae penetration. Cercariae move to adulthood and mate in the human body. Females lay eggs which are then re-released by urine or feces and the cycle starts again.

Water is the main risk factor. It is also the vehicle for disease transmission. The water network in Senegal is very dense and some fresh water points are permanent as in Senegalese river valley (Talla, Kongs and Verlé, 1992). This keeps a certain prevalence[2] even during the dry season. Climate, rainfall and irrigation schemes raise water point contact with human populations. Many situations such as poverty (lack of safe domestic water, sanitary infrastructure), and certain economic activities (agriculture, animal husbandry or fishery) lead to intense human-water contact. All these conditions lead to increasing ratio of people affected by the disease.

Therefore, schistosomiasis is in the boundary between both economic policy and hydrosphere management, and in the heart of potential conflicts between environmental protection and socio-economic constraints (Mayak manitu, 2001).

[2] Prevalence is a measure of the health of a population at a given time. For a given disease, it is calculated by dividing the number of cases of sick at a given time within a population by the total population.

The spread of schistosomiasis has serious consequences on economic development and children education suffers from this disease. To fight against this scourge, a National Program for the Fight against Schistosomiasis (PNLB) was established and attached to the Ministry of Health of Senegal. Other organizations, like the NGO Hope for Health (EPLS[3]), are actively working to produce vaccines against the disease. Although the goal of PNLB is to reduce morbidity and transmission, and to prevent seasonal expansion in areas of risk, it does not have a computer system capable of responding automatically and quickly to those needs. The slowness of data collection delays risk prediction. In addition, there is practically no numerical simulation model that can be used to make predictions. Therefore, the case of schistosomiasis in Senegal lends itself to a qualitative simulation for epidemiological monitoring.

Ontology scope definition

In order to define the ontology scope, we develop with experts the scenario "risks related to water" from which three use cases of ontology-based epidemiological monitoring process are specified.

Scenario: Risks related to water

Water is the vector of schistosomiasis transmission and the environment of parasite multiplication. Therefore, it is necessary to control potential sources that may raise presence of water points, or activities leading to contact between hosts (human here) and water points. In this scenario, monitoring focuses on rain. For example, the first rain marking the end of the dry season always presents a risk of emergence or spread of schistosomiasis in many parts of the country. Indeed, mollusks are able to sink into the ground and remain in a state of "diapause" when water points are dry. During this period, parasites can survive in human organism, as well as in mollusks. When the rain comes, water points are filled, mollusks come back to the surface and people go back to their activities. The parasites then, find all the necessary conditions for their reproduction and thus, the risk of disease emergence or spread increases.

Use case 1: Early prediction with lack of data

Taking into account rain occurrence, a qualitative simulation based on ontology reasoning can allow predicting its impact on disease prevalence in the future. Provisional measures can then be taken to prevent possible risks without more accurate data (amount of rainfall).

Use case 2: Support to numerical simulations

When the exact amount of rainfall and all needed data are collected, we can proceed with numerical simulation. In this use case, input simulation is the rainfall amount and analysis request can be to have disease prevalence at a given date. Three situations can occur:

- Situation 1: a simulation model *m* that has as input rainfall data and as output the schistosomiasis prevalence exists.
- Situation 2: no single simulation model – satisfying the request – exists, but composable models exist to satisfy user request. For example, composing a model *m1* giving mollusks density based on the rainfall and model *m2* given prevalence from mollusks density.
- Situation 3: no model composition is possible to satisfy user request. In this situation, the system can propose query refinements or make suggestions for an alternative scenario.

Modeling inputs, outputs, and objective functions of simulation models in the ontology, will help managing all these situations. For the first two situations, the monitoring support module (cf. figure 2) guides model selection and composition:

- Situation 1: The simulator loads and runs the model proposed by the monitoring support module.
- Situation 2: The monitoring support module specifies to the simulator the models that need to be composed and their arrangement. For example, the monitoring support module indicates to the simulator to select the

[3] http://www.espoir-sante.org/

model *m1* with rainfall data as input. The output (mollusk density) of this step is not returned to user as simulation is not yet finished. The monitoring support module then, indicates to the simulator to run the model *m2* with mollusk density as input. Let's note that composition is more complex than it seems, it may need simultaneous execution of composed simulation models.

It other situation, further data could be needed for simulation. In this case, the monitoring support module, based on ontology reasoning, can identify needed data and ask users to provide them. The monitoring support module can also guide interpretation of simulation results.

Use case 3: Guidance in implementation of action plans

An action plan consists of a sequence of actions. An example of action plan associated to rainfall is "to block access to some water points that could be potential sources of contamination". The action plan can be modeled as processes in the epidemiological monitoring ontology. Their occurrence conditions and effects need also to be modeled. Qualitative reasoning about processes particularly those dealing with actions related to risks, will guide specification of action plans and selection of the most appropriate ones.

An ontology for schistosomiasis monitoring

The content of an ontology is defined by its usage goals. In this application framework, we need to model domain and process knowledge to provide qualitative reasoning for an epidemiological monitoring system. Domain knowledge deals with schistosomiasis, risk factors and existing numerical simulation models. Process knowledge deals with monitoring, schistosomiasis spread, parasite life cycle, mollusk life cycle and the physiopathology of infected person. The structure presented below is defined with domain experts (cf. figure 3).

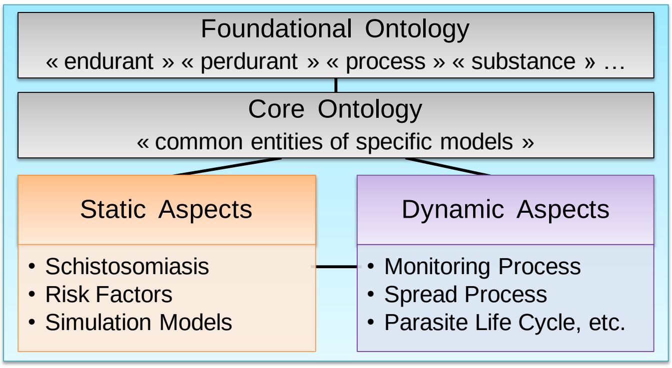


Figure 3. Ontology Structure for Schistosomiasis Monitoring

Domain knowledge (static aspects)

- *Schistosomiasis* is caused by a pathogen of the genus of trematodes. These genera belong to the family of *Schistosomatidae*. According to mollusk species living in fresh water in Senegal, only two parasites survive in the region: *S. Haematobium* and *S. Mansoni* that cause respectively urinary schistosomiasis and intestinal schistosomiasis. Figure 4 shows a simplified model of knowledge about the disease.

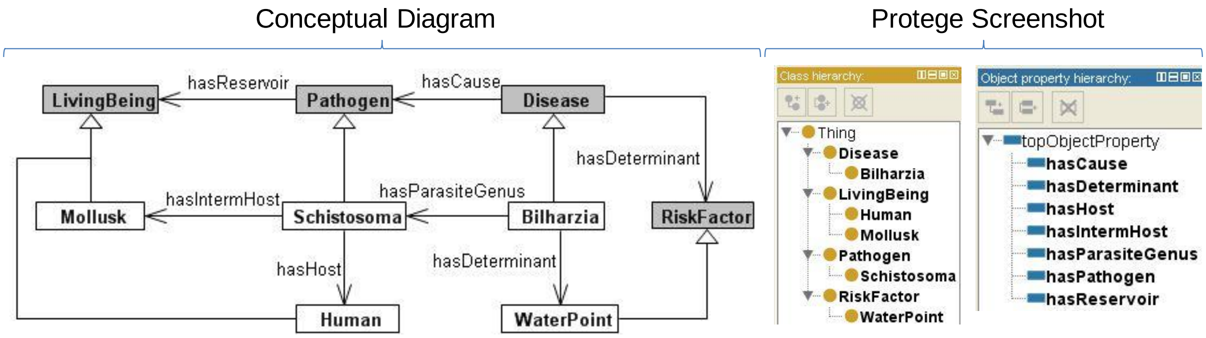


Figure 4. Schistosomiasis Conceptual Model

- *Risk factors* include all elements that could have an influence on disease spreading. Among these factors there are for example (cf. figure 5), biological factors, economical factors (agriculture, fishing), environmental factors (water, climate), logistic factors (hospitals, dams), behavioral factors (hygiene), etc.

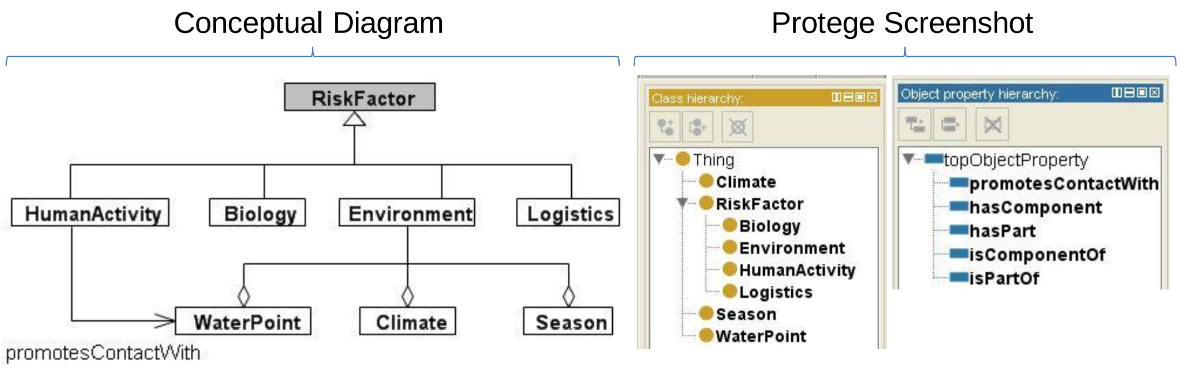


Figure 5. Conceptual Model of Schistosomiasis Risk Factors

- *Simulation models* allow numerical analysis of processes related to schistosomiasis (cf. figure 6). They allow reproducing some of the schistosomiasis dynamics with specific goals such as prevalence prediction or economic impact analysis. They can be described by their objective functions, inputs and outputs to facilitate their use and composition. Running simulation models aims to observe evolution – in time and space – of a part of a real system. Simulation models are static entities, while simulation executions reproducing dynamics of a phenomenon are processes.

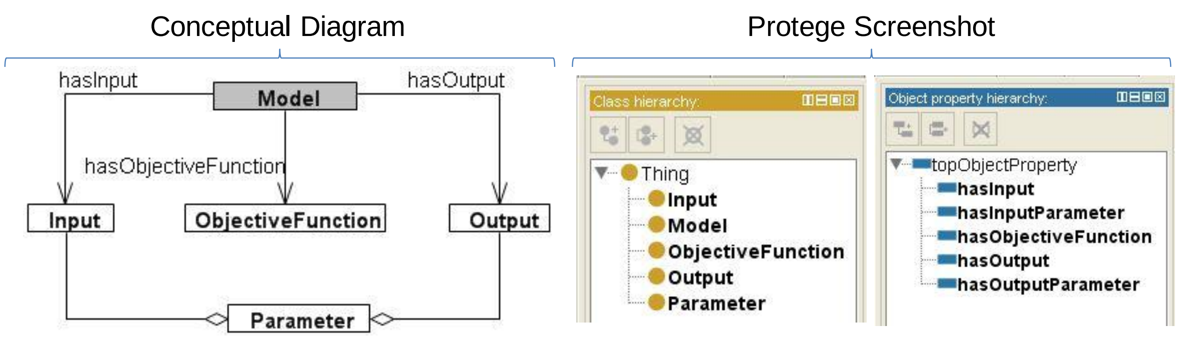


Figure 6. Conceptual Model of Simulation Models

Process knowledge (dynamic aspects)

Epidemiological monitoring can be modeled by two macro-wrapped processes: the monitoring process which aims to control the schistosomiasis spread process (cf. figure 7). Process Specification Language (Gruninger, 2009) is used here to formalize process specification.

The monitoring process includes several processes (cf. figure 7): (i) a surveillance process involving the processes of risk event detection and data collection, (ii) an analysis process, (iii) a decision process, and (iv) a decision assessment process.

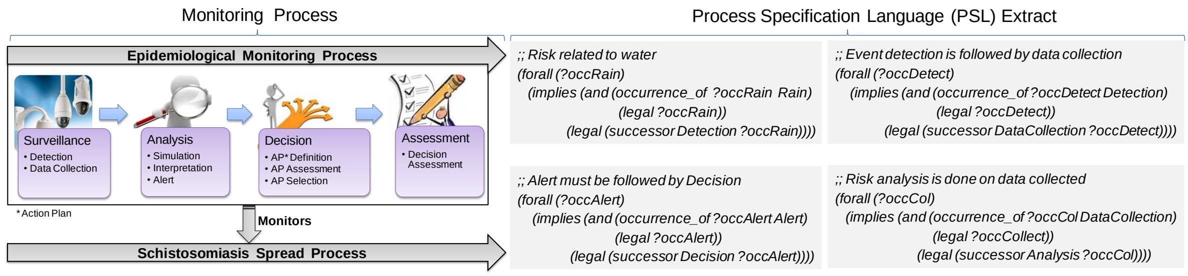


Figure 7. Epidemiological Monitoring Process

The analysis of schistosomiasis spread highlights a global dynamics involving several sub-processes and goals, a set of populations (humans, parasites and mollusks), and a set of organic (or biological) processes within individuals belonging to the different populations (human pathological reactions, life cycles of parasites and mollusks). Schistosomiasis spread process can be considered at two levels of granularity: *population level* (mainly the human population) and *individual level* (containing individual entities involved in the spread process). Individuals participate in the process at the population level and maintain processes at the organic level (cf. figure 8).

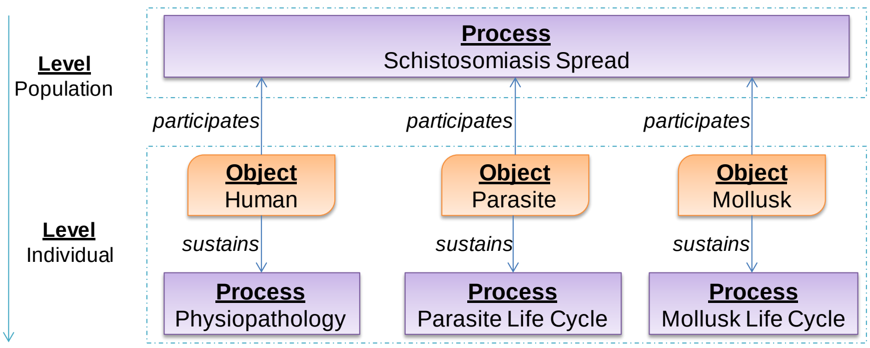
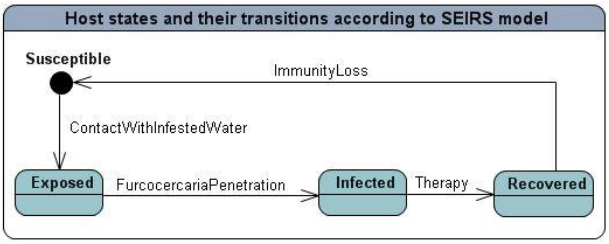


Figure 8. Multi-levels in Schistosomiasis Spread

In this section, we only present the modeling of the schistosomiasis spread process at population level. The modeling of organic processes is out of the scope of this paper. In epidemiology, there is a standard mathematical model representing infectious disease spread process at population level (Kermack and McKendrick, 1927). It is a compartmental model dividing a population into three groups of individual states: Susceptible, Infected and Recovered (SIR). For the schistosomiasis case study, Exposure (E) state is included between Susceptible and Infected states (SEIR). It is also possible for a recovered person to lose his immunity against the disease, the complete model is thus, SEIRS (cf. figure 9). The model below shows links between the different states of a host and their transitory processes.

State Machine Diagram



Process Specification Language (PSL) Extract

```
;; An occurrence of Infection means host was previously exposed
(forall (?occInfection)
   (implies (and (occurrence_of ?occInfection Infection)
                 (subactivity_occurrence ?occInfection ?occSpread))
            (exists (?occExposition)
               (and (occurrence_of ?occExposition Exposition)
                    (subactivity_occurrence ?occExposition ?occSpread)
                    (ealier ?occExposition ?occInfection)))))
```

Figure 9. Health States of Host and their Transitions According to SEIRS Model

CONCLUSION

This paper presents a new approach to epidemiological monitoring process implementation based on an ontology. The proposed ontology models both domain and process knowledge to facilitate knowledge sharing, communication between actors involved in the monitoring process, early predictions using qualitative simulations and decision-making support. The ontology-based approach is illustrated with a real application framework: schistosomiasis monitoring in Senegal. The contribution of the ontology is demonstrated through use cases showing how such an ontology could be used in the process of epidemiological monitoring for schistosomiasis in Senegal.

Currently, we are working on refining the built ontology particularly process modeling, to take into account multi-levels in disease spreading.

REFERENCES

1. Amouroux, E., Taillandier, P. and Drogoul, A. (2010) Complex environment representation in epidemiology ABM: application on H5N1 propagation. *in 'the 3rd International Conference on Theories and Applications of Computer Science.* Can Tho City, Vietnam.
2. Buton, F. (2006) De l'expertise scientifique à l'intelligence épidémiologique : l'activité de veille sanitaire, *Genèses*, 4, 65,71-91.
3. Drogoul, A., Treuil, J.-P. and Zucker, J.-D. (2008) Modélisation et simulation à base d'agents : Exemples commentés, outils informatiques et questions théoriques, Dunod, Paris.
4. Forbus, K. D. (1997) Qualitative Reasoning, The Computer Science and Engineering Handbook, 715-733.
5. Grenon, P. and Smith, B. (2004) SNAP and SPAN: Towards dynamic spatial ontology, *Spatial Cognition and Computation*, 4, 69-103.
6. Gruninger, M. (2009) Using the PSL Ontology, *In S. Staab, Handbook of Ontologies*, 419-431, Springer-Verlag, Berlin.
7. Kermack, W. O. and McKendrick, A. G. (1927) A Contribution to the Mathematical Theory of Epidemics, *Proceedings of The Royal Society of London. Series A, Containing Papers of A Mathematical and Physical Character*, 117, 772, 700-721, 1905-1934.
8. Le Moigne, J.-L. (1990) La modélisation des systèmes complexes, Dunod, Paris.
9. Lewis, D. (1986) On the Plurality of Worlds, Blackwell Publishers, Oxford.
10. MAYAKA MA-NITU, S. (2001) Etude epidemiologie de la bilharziose à schistosoma mansoni en milieu scolaire: Cas du groupement de Kiyanika. Thèse de doctorat en médecine, chirurgie et accouchement. Université Kongo.
11. Studer, R., Richard Benjamins, V. and Fensel, D. (1998) Knowledge engineering: Principles and methods. *IEEE Transactions on Data and Knowledge Engineering , 25*, 1-2, 161-197.
12. Talla, I., Kongs, A. and Verlé, P. (1992) Preliminary study of the prevalence of human schistosomiasis in Richard-Toll (the Senegal river basin), *Trans R Soc Trop Med Hyg*, 86, 2, 182-191.