A mortality representation for biomedical ontologies

Filipe Santana[1§], Fred Freitas[1],Roberta Fernandes[1], ZulmaMedeiros[2,3], Stefan Schulz[4,5],Daniel Schober[5]

[1]Informatics Center, Federal Universityof Pernambuco (CIn/UFPE), Recife, 50.740-560, Brazil;
[2]ParasitologyDepartment, AggeuMagalhães Research Center, Oswaldo Cruz Foundation, (CPqAM/Fiocruz), Recife, 50.670-420,Brazil;
[3]Pathology Department, Institute of Biological Sciences, University of Pernambuco, Recife, 50100-130, Brazil;
[4]Institute for Medical Informatics, Statistics, and Documentation, Medical University of Graz, 8036, Austria.
[5]Institute of Medical Biometry and Medical Informatics (IMBI), University Medical Center, Freiburg, 79104, Germany;

§Corresponding author

Email addresses:
FS: fss3@cin.ufpe.br
FF: fred@cin.ufpe.br
RF: rmf3@cin.ufpe.br
ZM: zulma.medeiros@cpqam.fiocruz.br
SS: stefan.schulz@medunigraz.at
DS: schober@imbi.uni-freiburg.de

Abstract

Background

Although many efforts are being devoted to the development of biomedical ontologies, very few, if any, encompass definitions of death. This concept and its related relations are bound to be used in many applications, e.g. data integration of national mortality and morbidity notification systems. In the current work, as part of the construction of the NTDO (Neglected Tropical Diseases Ontology), werepresented thecomplex processes andentities, characterized by theirtemporal milestones, the causality relations that hold among them, that relate to the representation of biological death. This effort supports the ongoing development of an application able to query simultaneously Brazilian morbidity and mortality databases, matching their registers when necessary.

Results

Several representational difficulties were overcome, mainly regarding to the concepts which are hard to be precisely grasped,*viz.* death and its relationships with injuries, diseases, processes, etc. We illustrate the ontology development process of the ontology parts related to death with four consecutive versions of the ontology. The purpose it this iterative development was stressing the typical problems encountered in representing complex biological events, as well as pointing out possible solutions. We designed the ontology together with its mortality parts following the principles of minimal encoding bias and minimal ontological commitment proposed by Tom Gruber, in order to foster the reuse possibilities by a wider range of applications.

Conclusion

The NTDO in its current status includes the mortality definitions described here, and allows for an accurate description of all of the processes related to diseases and injuries, including their evolution that ultimately can lead to death. Employing it together with other parts of NTDO, as the description of

pathogen transmission by arthropod vectors, enables building a detailed description of the processes and their relations, starting from the inoculation of a pathogen by a vector, until the death of an individual. The ontology with its mortality contents may serve many different aims, such as supporting tutoring systems, playing the role of a shared vocabulary in data integration solutions, etc.

Introduction

With the growing interest surrounding biomedical data available in large-scale experimental data sources and research results, *viz.* EMBL Nucleotide Sequence database (EMBL, http://www.ebi.ac.uk/embl/) and the Swiss-Prot Protein Database (Swiss-Prot, http://us.expasy.org/sprot/),many researchers are relying on ontologies for many tasks, such as information retrieval based on concepts (instead of terms), properly connecting the different senses of a concept which are related to data coming from these data sources, etc. Particularly, the use of ontologies facilitates the task of intelligent information integration [1]a great deal, since they assure interoperability in a safe way in terms of semantics of the concepts and sense disambiguation. Indeed, currently more than 250 ontologies are available in the BioPortal ontologies' library [2, 3]to represent a wide range of topics of biomedical interest and, also, to provide intelligent data retrieval from such databases.

Although many efforts are devoted to the development of ontologies, mainly following the structure, relations and annotations provided by the Gene Ontology [4] to integrate large data bodies, few others are focusing on the World Health Organization (WHO) working groups' needs and programs, such as the "Stop TB Partnership" program which aims at eliminate Tuberculosis from earth [5].

As the epidemiological surveillance actions aim at identifying and determining the factors which increase the spread a given disease, morbidity databases, such as the National Morbidity Notification Information System in Brazil (in Portuguese, *Sistema de Informação de Agravos de Notificação*, SINAN) [6], are used

as the main sources for deciding policies and concrete actions for disease prevention and control. Mortality databases are also related to the problem, once a piece of information stored in it, the main death cause, may be one of these diseases which receive interest from WHO. In Brazil, mortality data are stored in the Brazilian Mortality System (SIM - *Sistema de Informação de Mortalidade*). Both databases have a limited use to monitor health related events, like injuries or diseases, like TB [7].

Hence, on integrating the two databases, we are facing many interesting integration problems. For instance, we observed that the identifiers in both databases do not follow strict rules so as to prevent misidentification and leverage data integration. This type of problem can be considered as a syntactical one, which is usually dealt with algorithms that compute cumulative evidence from other pieces of the registers to decide for a matching. However, a more interesting semantic integration problem was also brought about while querying the two databases altogether: sometimes an individual dies due to a certain disease, but, instead of reporting that disease as the main cause of death, a typical symptom of it was annotated and carried to the database.

Therefore, a new requirement of the NTDO thus consists in encompassing the profile of an individual from life to death in case he had been struck by certain diseases, the so called Neglected Tropical Diseases. A main challenge for meeting this requirement resides in proposing a sound ontological representation of death. Many subtle aspects hamper a precise definition in this case: the conditions in which an individual is considered dead, the ontological problem of preserving identity of an individual or not, among others.

The purpose of the current study is to ontologically formalize foundational processes and other human lifecycle related events, particularly related to death, casualties, diseases and injuries, in the scope of the *Neglected Tropical Disease Ontology* (NTDO)[8, 9].

Ontologies, from a formal point of view, intend to describe the nature of entities from a given domain, ideally independently of possible linguistic variations of the terms referred by it. Accordingly, formal many ontologies are expressed by means of a formal semantics, when they are expressed in knowledge representation formalisms such as Description Logics (DL) [10], and its Web Ontology Language (OWL), which is a Semantic Web standard language recommended by the World Wide Web Consortium (W3C) [11].

The focus of the article is proposing a death representation which is able to overcome the problems posed by many of its subtleties, like preserving the identity meta-property[12], respecting cardinality constraints, etc. In order to serve as an example of iterative modeling, we outline here four versions for representing mortality, interleaving their presentation with a discussion of the representational problems or the complexity of reasoning arising from each. The consecutive versions and the iterative ontology construction process are explicited here with the intention of guiding other ontologists to detect and solve hidden problems with their biological representations.

Finally, in the end of the article we describe very briefly our use case on the integration of the morbidity and mortality databases described above. The death processes and related events incremented to the NTDO play a crucial role for the matching registers from the two databases.

Methods

For the representation of the new definitions, NTDO specialize classes and reuse relations provided by the upper level ontology BioTop[13], generalizing it upwards to the BioTop connection level and, sometimes, specializing downward to the required leaf node. Other classes and relations where imported from the General Formal Ontology [14, 15], mainly, for representing time intervals and its limits (boundaries). Throughout the text, we use the terms 'relation' and 'property' interchangeably, standing for binary relationships between two individuals, i.e. instances of a class.

Furthermore, the current work was based in well-established ontology construction guidelines available in Rector [16], which suggested the untangling of asserted graphs into disjoint orthogonal axes. We are also relying on the naming conventions provided by Schober et al. [17].

NTDO is built upon the Ontology Web Language 2 (OWL2)[11], by means of the ontology editor Protégé v.4.1, which uses the reasoned HermiT[18].

Apart from the literature review that helped us on creating the ontology contents related to death, other relevant sources were the Brazilian Morbidity and Mortality systems themselves[6, 19].Thereunto, at some extent, we grounded our definitions on the way death cases are reported in the systems' databases[19].In the mortality system, cases include "the main cause of death" as well as other secondary causes, which can be directly associated with certain diseases.

From the epidemiological point of view, these can only be completely defined *a posteriori*, since a previous cause (illness / injury) can only be linked to the main cause of death in a *post mortem* analysis (by autopsy, for instance), or the statementfroma forensic medicine service or the physician that took care of the patient until her death. Another relevantaspect, from a conceptual standpoint, is the fact that the

axioms we defined here mention only causal relationships. However, the notion of causality used and necessary for the representation is based on the observer of the process, *i.e.* the health professional (physician) who certified the cause of death.Taking as an example a death record in a mortality notification database, *viz.* SIM, a physician certifies the underlying cause of death and sometimes secondary ones.

Results

In this section, we describe the ontological representation of mortality. The models proposed here are intended to cover the representation of death occurring by most means, although we are focused on the ones related to diseases as their primary causes. Thus, the representations describe the processes and events,starting from the transmission of a pathogen, going throughthe disease contamination, and finally arriving at a possible passing. In the next subsections, we will discuss and represent many complex issues and aspects of the modeling.

Representing Injury and Death

Since the terms 'birth', and 'death'are dual, both referring to the term 'life', and no standard definition of them is absolutely accepted in the literature, it is necessary to introduce the notion of them to proceed to more complex considerations. First of all, the representation of death presented here is based on the lifecycle displayed in Figure 1 below.

Taking the birth as the starting point, an individual lifespan continues until her death, and necessarilyincludes the beginning of 'biological death processes' (if a register in the mortality database refers to her), one or more of which will ultimately lead to her death. That definition is grounded on the widely known fact that several factors can influence the lifetime of an individual, such as an accident or an illnessthat provokes her death. Many mortality systems worldwide assume that only a single main cause of death is possible, although this is more of a notification design decision that helps build health statistics. On our ontology, we follow this stance; nevertheless extending it to accommodate many causes is a quite straightforward process that causes no additional computational costs for the reasoned that will deal with the queries.

Firstly, 'life' is thought of in this work as described by Koshland [20]. It relates the inherent capability of being naturally programmed (as described in each individual's DNA), as well as of accomplishing kinetic interactions (such as genome and metabolism), capable of mutating and enabling selection of individuals. The bearer of survival skills is a body compartmentalized in cells and / or organs, which can metabolize substances to generate energy for adaptation, regeneration and reproduction.

The definition of 'Birth' presumed in this work is also based on the descriptionof "live birth" provided by theBrazilian Institute of Geography and Statistics (*InstitutoBrasileiro de Geografia e Estatistica* –IBGE) [21]. It corresponds to the complete expulsion or extraction of a product generated by the maternal body after conception, which after separated from the maternal body, breathe or give some other signs of life, *e.g.* heartbeat, voluntary muscle contraction, umbilical cord contraction, even ifthe cord was cut or not, and whether or not the placenta was expelled. Conversely, "death" means absence of brain functions and cessation of all biological functions, inherent to the human body [22].

At a given moment in the life of an individual, she can acquire, for instance, by vector transmission, a certain disease, *e.g.*dengue fever, which may causeher to die earlier than expected, depending upon the disease type and herhealth conditions. Thedeathof an individual is usually linked to one or more causes, which are determined dependent upon medical or medical-legal investigation. In medical terms, the cause of a death is a dysfunction of the physiological state of the individual which leads to death. It can be related to previous biological death processes, registered or not, in the life of an individual.

Thus, it is possible to describe and trace the sequence (or overlapping) of the deleterious processes which pervaded the life of that individual, from her birth to her death. Some of them may damage the overall physiological state of the organism so as to initiate a physiological deathprocess, leading to death itself. This sequence of events is sometimes evidenced by the records of an individual when the cause of death was previously registered in a morbidity system, *i.e.* the cause of death was already known.

However, there are major difficulties related to the accurate representation of the processes that makes an individual die, mainly due to the complexity in relating sequences of processes and time, with a precisedescription of when each process took place. Nevertheless, this exact information is probably not important at all; instead what counts is the knowledge of the causal and temporal relations of symptoms in a disease when they are detected (by the registry in a morbidity database), because the time constraints involving them, e.g. during a TB, a cough with secretion is followed by a pulmonary infection, can be checked in morbidity and mortality notifications.Although the symptoms are not usually found out in their beginning, we assume as premise that they for our aims they exist when they are notified for the morbidity information system. We also presume the notion of 'event' provided by GFO, as a fact that occurs in an instant. We use this concept to state the relation among the behavior the patient exhibits with some processes which are temporarily limited (right or left time boundaries) by events[15].

Next, we present the main challenges related to the representation of entities and conceptualizations introduced until now. For this purpose, the representation is divided into four versions, which demonstrate the evolution of the model to a final proposal.

The Mortality Models

Several challenges were encountered in creating a coherent representation for a mortality event, such aspreserving the identity of related individuals by setting cardinalities,and representing the resulting ontology in a decidable DL, thus posing no difficulties for the reasoners that will solve its user queries. Each of these items is discussed in the consecutive versions until we arrive at a satisfactory model.

    ✓ Version 1

An initial definition of death could be:

ntdo:*DeathEvent*equivalentTogfo:*Event*

   and (biotop:**hasLocus** some ntdo:*GeographicLocation*)

   and (biotop:**hasPatient** some *DeadOrganism*)   (1)

   and (biotop:**precededBy** some ntdo:*BiologicalDeathProcess*)

   and (ntdo:**hasInstant** some *PointInTime*)

indicating that an event of death is an event that occurs in at least one location, and in which a dead organism is a participant. It also states thatthere is one or more biological death processes as part of the death event and it takes place at a given moment in time.

✓ Analysis

Despite this version grasps the general idea of connecting death with biological processes, such definition brings several imprecisions, mainly on how to preserve identity and the relation of identity andcardinalities.

First, the purpose of representing this class is to convey information about the death of a single organism. However, the axiom expresses no cardinality constraint, what may give rise to the unintended interpretation (according to the mortality database assumption) if more than an individual dying at once in a same death event. Moreover, according to the class definition, there is no guarantee that the living and the dead body are identical, since the patients of the *DeathEvent* and *BiologicalDeathProcess* may not be the same. Nothing in the axiom states that the *BiologicalDeathProcess* relates to individual involved in the relation **hasPatient**.

Besides, the idea of a living organism that, at some point in time, is transformed into a dead one causes further representational problems. First, it is incompatible with BioTop, which classifies only living

individuals. As a consequence, according to BioTop, a dead human correctly shall not be considered a human any more, although possessing human organs, features, etc. Besides losing its "humanity", identity is lost too, since any classification of living beings is rigid [12], so if an individual is not an instance of a rigid class anymore, it ceases its existence. Even if we assume that this description corresponds to the anti-rigid class *phased sortal*[12], *i.e.* entities which change phase (from "living" to "dead"), it is not clear until when identity should be preserved: the ashes of a dead organism should be identified as the dead one?

✓   Version 2

A solution to circumvent the representational problems of version 1, and, not surprisingly, the common choice of ontologists who modeled all other biological ontologies found in the literature, is simply not representing the*DeadOrganism*class, the entity which causes the whole confusion. It indeed plays no role in most health applications. The new solution then consists in representing*LivingOrganisms* based on their temporal existence, limited by two time points, as described in GFO [15].

Such representationemploys the definition of gfo:*Chronoids,* as entities *sui generis*, i.e. not defined as a set of points, but as intervals bounded by very small intervals, thus considering time represented as a continuum. Thus, every *Chronoid* has two outer boundaries, known as time limits (gfo: *TimeBoundary*) or points in time. In GFO, there are two types of time boudarieswhich represent the right (start) and left (end) temporal limits, the gfo:*LeftTimeBoundary* and the gfo:*RightTimeBoundary*. By definition, they cannot assume the same values in a single *Chronoid* [15]. A schematic representation can be found figure 2.

For the sake of clarity, we show here the GFO definitions of *Chronoid* and its time boundaries:

gfo:*Chronoid* subClassOf

(gfo:**hasLeftTimeBoundary** exactly 1 gfo:*LeftTimeBoundary*) and   (2)

(gfo:**hasRightTimeBoundary** exactly 1 gfo:*RightTimeBoundary*)

gfo:*LeftTimeBoundary* equivalentTo gfo:*TimeBoundary*   (3)

  and (gfo:**leftTimeBoundaryOf**some gfo:*Chronoid*)

gfo:*LeftTimeBoundary* equivalenTo gfo:*TimeBoundary*   (4)

  and (gfo:**leftTimeBoundaryOf** some gfo:*Chronoid*)

Chronoids are temporarily related according to James Allen's interval temporal algebra[23], which considers the relation of two chronoids in one of the situations depicted in figure 3.

When there are *Chronoid*s in sequence, the right time limit of a preceding process must be contiguous with the left of the subsequent, represents the beginning of a new *Chronoid* and the end of the previous. It is worth noting that the mereological sum of *Chronoid*s represents the notion of temporal region [15]. Following the GFO perspective, entities that exist entirely in a time interval are referred to as gfo:*Presential*[15]. This class includes material entities that are represented in BioTop as biotop:*MaterialEntity*. Thus, the former can be regarded as a superclass of the second. From this assumption, the axioms below should be included:

biotop:*MaterialEntity* subClassOf *Presential*   (5)

biotop:*LivingOrganism* subClassOf *Presential*   (6)

stating that there is only one time interval corresponding to the existence of a living organism (its lifespan).

Processes are projected (GFO: **projectsTo**) to *Chronoids*[15]. In order to avoid mismatches in NTDO, the class gfo:*Process* must be mapped to the class biotop:*ProcessualEntity*.

Finally, the *DeathEvent* should be modified to replace a *DeadOrganism* by a *LivingOrganism*, as follows:

ntdo:*DeathEvent* equivalentTo gfo:*Event*

      and (biotop:**hasLocus** some ntdo:*GeographicLocation*)

      and (biotop:**hasPatient** some biotop:*LivingOrganism*)   (7)

      and (biotop:**precededBy** some ntdo:*BiologicalDeathProcess*)

      and (ntdo:**hasInstant** some ntdo:*PointInTime*)

✓ Analysis

On the one hand, the ontological problems with the existence of *DeadOrganism*s are solved, including the identity metaproperty problem. Indeed, in this version, instances of *LivingOrganism* are formed at a certain time point (gfo:*LeftTimeBoundary*) and destroyed in another (gfo:*RightTimeBoundary*). On the other hand, by definition the relationship biotop:**hasPatient** allows more than one element in the range, which can lead to the erroneous interpretation that an event of deathmayhappen to several people simultaneously.

Moreover, it still contains three identity problems:

(a) The one between the *DeathEvent* and the *BiologicalDeathProcess* patients, which still persists;

(b) the set of definitions stated up to that point neither include the moment of death nor make it identical to the end of the *BiologicalDeathProcess* that led to it*;* and

(c) the same applies to the dying *LivingOrganism*, whose *RightTimeBoundary* should coincide with both the *DeathEvent* and the end of the *BiologicalDeathProcess* that arrived at it*.*

✓ Version 3

The last mentioned problem can only be solved with the representation of instantaneous events. For this purpose, the class gfo:*Event* was reused, as being a process milestone that takes place in the beginning or end of a process.

We now need the definition of a relation to ascribe the exact instant when a death takes place (again considered as reported in the mortality database). This can be reached by defining the relation **hasInstant** asthe right border (the end) of the *BiologicalDeathProcess* related to the *DeathEvent*, as below:

ntdo:**hasInstant** = gfo:**projectsTo** o gfo:**hasRightTimeBoundary** (8)

However, in order to solve identity problem between the *DeathEvent* and the *BiologicalDeathProcess* patients from version 2, the DL agreement operator ($\doteq$) will come into play. This operator is used in chains of properties to indicate that the instances to be described are connected. It is worth stressing, the difference between the two operators, $\doteq$ and =. The former represents a coincidence in the value of two properties, or, in other words, a reference to a very same object, whilst the latter defines a formation rule for a relation.

The new definition of a *DeathEvent* goes below:

ntdo:*DeathEvent* equivalentTo ntdo:*Event*
     and (biotop:**hasLocus** some ntdo:*GeographicLocation*)
     and (ntdo:**hasPatient** some biotop:*LivingOrganism*)
     and (ntdo:**precededBy** exactly 1 biotop:*BiologicalDeathProcess*)
     and (ntdo:**hasDeathPrimaryCause** exactly 1 biotop:*BiologicalDeathProcess*) (9)
     and (ntdo:**hasDeathPrimaryCause** $\doteq$ ntdo:**precededBy**)
     and (ntdo:**hasPatient** $\doteq$ ntdo:**precededBy** o ntdo:**hasPatient**)

and (ntdo:**hasInstant** $\doteq$ ntdo:**precededBy** o gfo:**hasRightTimeBoundary**)

It defines:

-where a death event takes place, (biotop:**hasLocus** some ntdo:*GeographicLocation*),

- which deceased organism is its patient (ntdo:**hasPatient** some biotop:*LivingOrganism*),

- which biological death process preceded it(ntdo:**precededBy** exactly 1 biotop:*BiologicalDeathProcess*), and

- whichprocessual entity was the death primary cause (ntdo:**hasDeathPrimaryCause** exactly 1 biotop:*BiologicalDeathProcess*).

The agreement conditionsare the more important definitions, given that they solve the last version's shortcomings. They ensure that:

-the death occurs exactly when the *BiologicalDeathProcess* is finished

(ntdo:**hasInstant** $\doteq$ ntdo:**precededBy** o gfo:**hasRightTimeBoundary**),

- that the biological death process that caused the death precedes it (ntdo:**hasDeathPrimaryCause** $\doteq$ ntdo:**precededBy**);and

-that a deceased person is the same who participated in the injury event that led to the death, thus retaining the identity of the patient (ntdo:**hasPatient**$\doteq$ntdo:**precededBy**ontdo:**hasPatient)**.

The modeling practice using in the last condition that refers to patients is known by the DL practitioners as left identity[24].

Completing the task of modeling mortality, the class ntdo:*BiologicalDeathProcess* was created to indicate the existence of (not completely known) processes that occur to the dying organism, which collaborate to the death. A biological death process (from disease to death) is therefore a biological processual

entitywhich is caused by an injury or other biological processes (but, of course, notcaused by biological

death processesthemselves). It has as patient a convalescent organismand its duration is delimited:

ntdo:*BiologicalDeathProcess* equivalentTo biotop:*BiologicalProcessualEntity*

  and (biotop:**causedBy** exactly 1 (ntdo:*InjuryEvent* or

    (biotop:*BiologicalProcessualEntity*   (10)

      and (not (ntdo:*BiologicalDeathProcess*)))))

  and (ntdo:**hasPatient** some biotop:*LivingOrganism*)

This axiom addresses the processes that occur prior to the death event and after an injury event or disease.
It is worth mentioning two facts about *BiologicalDeathProcesses*:

  -oncetheyare*BiologicalProcessualEntit*ies*,* they are projected into *Chronoids,* and thus exist in
time intervals limited by *TimeBoundaries*; and

  - more than one *BiologicalDeathProcesses*can be assigned to a same individual, but only one will
eventually be realized when the death cause is identified by a physician or forensic professional.

Finally, it is important to disambiguate it from an *InjuryEvent*. For the latter, it is necessary to
determinewhere geographically it took place (ntdo:**hasGeographicLocation**), its cause and the injured
patient. Injury causes are described here as any other process which affect patients, but notbiological in
nature. All of this is ascribed in the axiom below:

ntdo:*InjuryEvent* equivalentTo ntdo:*Event*

  and (ntdo:**hasGeographicLocation** some ntdo:*GeographicLocation*)

  and (biotop:**causedBy** some (biotop:*ProcessualEntity*   (11)

    and (not biotop:*BiologicalProcessualEntity*)))

  and (ntdo:**hasInjuriedPatient** some biotop:*LivingOrganism*)

Despite not being the focus of the current work, which is about deaths caused by diseases, it is necessary to distinguish diseases and disorders, once a distinction is already stated in BioTop[25]. A disorder is caused by an accident, a lesion, or a fracture and can lead to a disease. Thus, disorders follow injuries.

✓ Analysis

For the purpose of the modeling, our goalsseem to be fulfilled, given that no identity problems are left. However, a hidden problem not related to the representation but to the reasoning: if equalities are not built over property chains of functional properties, i.e. the relations which permits only one instance in its range, in role-value-maps (like role compositions or agreements), then inference becomes undecidable[26].

✓ Version 4

Those problems are related to the cardinality of the relationship biotop:**hasPatient**. For our purposes, this relation must be functional, *i.e.* each element of the domain must be mapped to at most one element of the range, and unfortunately BioTop does not define it as such. Therefore, in order to meet this requirement, we created the following subproperties, all functional:

Functional (**hasDeathPatient, hasConvalescentPatient, hasInjuredPatient**) (12)

indicating that an injury or death event admits only one instance of patient (*i.e.* person). For instance, the functional property biotop:**hasInjuredPatient** fits perfectly to most healthcare notifications, since it refers exclusively to a single person. The property **hasConvalescentPatient** is only employed in *BiologicalDeathProcesses,* while analogously **hasDeathPatient** is used in the definition of the *DeathEvent*(Fig 4.)*,*

as can be seen below:

ntdo:*DeathEvent* equivalentTo ntdo:*Event*

    and (biotop:**hasLocus** some ntdo:*GeographicLocation*)

    and (ntdo:**<u>hasDeathPatient</u>** some biotop:*LivingOrganism*)   (13)

    and (ntdo:**precededBy** exactly 1 biotop:*BiologicalDeathProcess*)

    and (ntdo:**hasInstant** $\doteq$ ntdo:**precededBy** o gfo:**hasRightTimeBoundary**)

    and (ntdo:**hasDeathPrimaryCause** exactly 1biotop:*ProcessualEntity*)

    and (ntdo:**<u>hasDeathPatient</u>** $\doteq$ ntdo:**precededBy** o ntdo:**<u>hasConvalescentPatient</u>**)

This new definition has the advantage of stressing explicitly the fact that the death patient coincides with the *BiologicalDeathProcess* convalescent patient.

For a better understanding, a schematic model highlighting the main classes, relations and mappings presented in the axioms for representing deathis depictedbelow (Fig. 5).

It deploys the transitional profile of a biotop:*LivingOrganism*, from life to death. It also shows some agreements required to express the temporal sequence of processes.

Discussion

Since no ontology on mortality is available, we will accomplish a comparison with other works that discuss mortality epistemologically. A related work about a discussion onontology of death was conducted by Thomasma [27], but without asserting a definition for death in logical terms. His work enlists related terms and provides some connections among them instead.

Another argument supporting our point of view, was described by Schrader [28]. His work indicates that Immanuel Kant reported the fact that indistinguishable entities (like 'death') can be distinguished by a particular and unique situation against 'space' and 'time'. Such standpoint is fulfilled by the model shown here.

Conclusions

In the current work, as part of the construction of the NTDO (Neglected Tropical Diseases Ontology), we represented the complex processes and entities, characterized by their temporal milestones, the disease causal relations that hold among them, (according to the data notification rules used for the Brazilian morbidity and mortality information systems, which also applies to many other countries worldwide) that relate to the representation of biological death.

Several representational difficulties were faced, mainly regarding to the concepts which are hard to be precisely grasped, *viz.* death and its relationships with injuries, diseases, processes, etc. The *step-by-step* modeling description, displayed herein the four consecutive versions of the ontology aim at stressing the typical problems encountered in representing such complex biological events, like preserving identity of the participant entities, assigning the correct cardinalities for each relation in a defining axiom, and, last but not least, decidability issues and computational overheads linked to each new DL constructor set introduced with the representation. The analysis of the problems present in each version motivates the next version as a possible solution to such problems.

The NTDO in its current status allows for an accurate description of all of the processesrelated to diseases and injuries, including their evolution that ultimately can lead to death. Using it together without her parts of the ontology, such as the description of pathogen transmission by arthropod vectors (present in[8]), a rich description of the biological processes and their subtle relationships in neglected tropical diseases, like the dengue and yellow fevers, lymphatic filariasis, and others, can be described in detail, starting from the inoculation of a pathogen by a vector, until the death of the infected individual. Therefore, the ontology, with the current addition of mortality related contents, is ready to serve many different

purposes, such as supporting tutor systems, playing the role of a shared vocabulary in data integration solutions, etc.

Currently, we are working in a use case that matches morbidity and mortality databases. The ontology is being used for many purposes. Once the process of populating the two databases does not include checking whether the notified data is correct against the constraints imposed by the complex axioms (such as impossibility of a certain disease occur in some areas) and rectifying wrong data (such as symptoms of a disease mistakenly considered as main causes of death instead of the disease itself), carefully crafted axiom-rich biomedical ontologies such as NTDO can be used for finding errors in the databases as well as monitoring which stages of the diseases are more prevalent.

We designed the ontology together with its mortality parts following the principles of minimal encoding bias and minimal ontological commitment [29] in order to foster the possibilities of reuse by a wide range of applications. If the goal is to perform integrated queries over the two or more databases at the same time, ontologies can play an important role in promoting data and information retrieval, data integration, data exchange and semantic interoperability in order to support policies' decision making.

Our use case above mentioned consists of an application capable of The ontology helps in tasks typically related to reasoning, such as checking whether the notified data is correct against the constraints imposed by the complex axioms (such as impossibility of a certain disease occur in some areas) and rectifying wrong data (such as symptoms of a disease mistakenly considered as main causes of death instead of the disease itself). These tasks could not be carried out without an axiom-rich ontology at hand.

Once morbidity and mortality databases do contain registers for the same person, in case the death causes of a given individual were previously identified, both registers shall be semantically related and a causal

profile can be stated. Those connections are explicitly represented in the ontology, without binding the former to any process of notification via system, in order to comply with the principles of minimal encoding bias and minimal ontological commitment [29]. This design decision was taken for fostering its reuse by other applications.

Competing interests

In the past five years, none of the authors received reimbursements, fees, funding, or salary from any organization that may, in any way, gain or lose financially from the publication of this manuscript, either now or in the future. The organizations which supportedthe German participants are:

- Deutsche Forschungsgemeinschaft (German Office of Research);
- and the German Ministry of Education and Research.

Noneof the authors hold stocks or shares in an organization that may gain or lose financially from the publication of this manuscript, either now or in the future.

None of the authors is currently applying for any patents relating to the content of the manuscript.

None of the authors has other financial competing interest.

None of the authors has other non-financial competing interest related to the manuscript.

Authors' contributions

FS – Main idea about the article, development and review of most part of the manuscript;

FF – Main idea about the article, development and review of most part of the manuscript;

RF – Development of the ideas concerning the linkage between morbidity and mortality data;

ZM – Supported the development of the ideas, mainly the parts which concerns to epidemiological issues;

SS – Reviewed the manuscript and contributed with ideas concerning the article's content;

DS – Reviewed the manuscript and contributed with ideas concerning the article's content.

Acknowledgements

Reference

1. Mena E, Kashyap V, Illarramendi A, Sheth A: **Domain Specific Ontologies for Semantic Information Brokering on the Global Information Infrastructure 1 Introduction**. In *Formal Ontology in Information Systems*. IOS Press; 1998.

2. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey M-A, Chute CG, Musen M a: **BioPortal: ontologies and integrated data resources at the click of a mouse.** *Nucl. Acids Res.* 2009, **37**:W170-3.

3. **BioPortal** [http://bioportal.bioontology.org/].

4. The Gene Ontology Consortium: **Gene Ontology : tool for the unification of biology**. *Nature Genetics* 2000, **25**:25-29.

5. Stop TB Partnership: *The Global Plan to Stop TB 2011-2015*. 2011:101.

6. **Sistema de Informação de Agravos de Notificação (SINAN)** [http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21383].

7. Selig L, Belo M, Cunha AJLA da, Teixeira EG, Brito R, Luna AL, Trajman A: **Óbitos atribuídos à tuberculose no Estado do Rio de Janeiro**. *Jornal Brasileiro de Pneumologia* 2004, **30**:335-342.

8. Santana F, Schober D, Medeiros Z, Freitas F, Schulz S: **Ontology patterns for tabular representations of biomedical knowledge on neglected tropical diseases**. *Bioinformatics* 2011, **27**:i349-i356.

9. **Neglected Tropical Disease Ontology** [http://www.cin.ufpe.br/~ntdo/].

10. Baader F, McGuinness DL, Nardi D, Patel-Schneider P: *The Description Logic Handbook: Theory, implementation, and applications*. 1st edition. Cambridge: Cambridge University Press; 2003.

11. **OWL 2 Web Ontology: Language Document Overview** [http://www.w3.org/TR/2009/REC-owl2-overview-20091027/].

12. Guarino N, Welty C: **A Formal Ontology of Properties**. In *Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management*. edited by Dieng R, Corby O Menlo Park: AAAI Press; 2000.

13. Beisswanger E, Schulz S, Stenzhorn H, Hahn U: **BIOTOP : An Upper Domain Ontology for the Life Sciences**. *Applied Ontology* 2008, **3**:205-212.

14. Heller B, Herre H: **Ontological Categories in GOL**. *Axiomathes* 2004, **14**:57-76.

15. Herre H, Heller B, Burek P, Hoehndorf R, Loebe F, Michalek H: *General Formal Ontology (GFO): A Foundational Ontology Integrating Objects and Processes. Part I: Basic Principles*. Leipzig: 2007:85.

16. Rector AL: **Modularisation of domain ontologies implemented in description logics and related formalisms including OWL**. In *Proceedings of the international conference on Knowledge capture - K-CAP ʼ03*. New York, New York, USA: ACM Press; 2003:121.

17. Schober D, Smith B, Lewis SE, Kusnierczyk W, Lomax J, Mungall C, Taylor CF, Rocca-Serra P, Sansone S-A: **Survey-based naming conventions for use in OBO Foundry ontology development.***BMC bioinformatics* 2009, **10**:125.

18. Motik B, Shearer R, Horrocks I: **Hypertableau Reasoning for Description Logics**. *Journal of Artificial Intelligence Research* 2009, **36**:165-228.

19. **Sistema de Informação de Mortalidade (SIM)** [http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21377].

20. Koshland DE: **The seven pillars of life.***Science (New York, N.Y.)* 2002, **295**:2215-6.

21. **Nascido Vivo** [http://www.ibge.gov.br/home/estatistica/populacao/registrocivil/nascido_vivo.shtm].

22. Miller FG, Truog RD: **Decapitation and the definition of death.***Journal of medical ethics* 2010, **36**:632-4.

23. Allen JF: **Maintaining knowledge about temporal intervals**. *Communications of the ACM* 1983, **26**:832-843.

24. Baader F, Schulz S, Spackman K, Suntisrivaraporn B: **How should parthood relations be expressed in SNOMED CT?** In *Ontologien in Biomedizin und Lebenswissenschaften (OBML) 2009*. edited by Herre H, Kelso J, Schulz S Leipzig: 2009.

25. Schulz S, Spackman K, James A, Cocos C, Boeker M: **Scalable representations of diseases in biomedical ontologies.***Journal of biomedical semantics* 2011, **2 Suppl 2**:S6.

26. Schmidt-Schauss M: *Computational aspects of an order-sorted logic with term declarations*. Lecture no. Berlin; New York: Springer Verlag; 1989:171.

27. Thomasma DC: **The comatose patient, the ontology of death, and the decision to stop treatment**. *Theoretical Medicine* 1984, **5**:181-196.

28. Schrader GA: **Heidegger's Ontology of Human Existence**. *The Review of Metaphysics* 1956, **10**:35-56.

29. Gruber T: **A translation approach to portable ontology specifications**. *Knowledge Acquisition* 1993, **5**:199-220.

30. **USDL Foundation Module as RDF** [http://people.kmi.open.ac.uk/carlos/174].

Figures

FIGURE 1– The life cycle of a human, including its main processes and the time regions in which they occur.

(PIC 1 File)

FIGURE 2 – The interval described for a gfo:*Chronoid* and its limits (gfo:*LeftTimeBoundary* and gfo:*RightTimeBoundary*) in a temporal axis.

(PIC 2 File)

FIGURE 3 – Graphical description of James Allen's interval temporal algebra [30 *apud* 23].

(PIC 3 File)

FIGURE 4 – The participants of *InjuryEvent*, *DeathEvent* and *BiologicalDeathProcess*.

(PIC 4 File)

FIGURE 5 – Main classes and relations used to model a *DeathEvent*.

(PIC 5 File)