

# **Towards an ontological representation of morbidity and mortality in Description Logics**

Filipe Santana<sup>1§</sup>, Roberta Fernandes<sup>1\*</sup>, Daniel Schober<sup>2\*</sup>, Stefan Schulz<sup>2,4\*</sup>, Zulma Medeiros<sup>3,5\*</sup>,  
Fred Freitas<sup>1\*</sup>

<sup>1</sup>Informatics Center, Federal University of Pernambuco (CIn/UFPE), Recife, Brazil;

<sup>2</sup>Institute of Medical Biometry and Medical Informatics (IMBI), University Medical Center,  
Freiburg, Germany;

<sup>3</sup>Parasitology Department, Aggeu Magalhães Research Center, Oswaldo Cruz Foundation,  
(CPqAM/Fiocruz), Recife, Brazil;

<sup>4</sup> Institute for Medical Informatics, Statistics, and Documentation (IMI), Medical University of  
Graz, Austria.

<sup>5</sup>Pathology Department, Institute of Biological Sciences, University of Pernambuco, Recife,  
Brazil;

\*These authors contributed equally to this work

§Corresponding author

Email addresses:

FS: fss3@cin.ufpe.br

RF: rmf3@cin.ufpe.br

DS: schober@imbi.uni-freiburg.de

ZM: zulma.medeiros@cpqam.fiocruz.br

SS: stefan.schulz@medunigraz.at

## Abstract

**Background.** Although many tackle the development of biomedical ontologies, very few encompass sound definitions of death. Nevertheless, this concept has its relevance in epidemiologic applications, e.g. data integration within mortality notification systems. We here introduce an ontologic representation of the complex biological qualities and processes that inhere in organisms transitioning from life to death. We further characterize them by causal processes and their temporal borders.

**Results.** 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. We illustrate an iterative optimization of definitions within four versions of the ontology, so as to stress the typical problems encountered in representing complex biological processes as well as pointing out possible solutions, focusing on the interdependencies and need to change further parts if one part is changed.

**Conclusion.** Our presented axiomatic mortality definitions allow for an accurate description of biologic processes related to the transition from healthy to diseased, injured, and up to a final death state. Exploiting such definitions embedded into NTDO based descriptions of pathogen transmissions by arthropod vectors, the complete sequence of infection and disease processes can be described, starting from the inoculation of a pathogen by a vector, until the death of an individual, preserving identity of the patient. An example usage of the mortality representation relies in the impossibility of assuring the cause of death due to a certain cause (e.g. a neglected tropical disease like Dengue Fever, Chagas Disease, among others), and rectification of wrong

data, like indirectly related disease causes (e.g. heart attack of a patient with Chagas disease) mistakenly considered as main causes of death instead of the disease itself.

## **Introduction**

With the growing need to cope with large-scale biomedical data researchers have been relying on ontologies for aligning to a common, shared and computer-interpretable meaning for linguistic terms describing such data. As the formality and defined semantics of the ontology representation languages allow for detection of logical equivalences in heterogeneously represented descriptors, they can be used for the task of intelligent information integration and interoperability [1]. Indeed, more than 250 ontologies<sup>1</sup> are available in the BioPortal ontology library [2] for representing different topics of biomedical interest and, also, to ultimately foster intelligent data retrieval from databases using these ontologies to annotate their resources.

Although many efforts are devoted to the development of genomics and metabolomics ontologies, often motivated by the role model Gene Ontology [3], few are focusing on patient and disease centered data. This is required, for instance, in epidemiology to study the dynamics of diseases, and to define health policies for epidemiological surveillance.

As the epidemiological surveillance actions aim at identifying factors which promote the development of diseases, morbidity databases are used as the main sources for disease prevention and control, such as the National Morbidity Notification Information System in Brazil (SINAN) [4]. Furthermore, mortality databases describing the cause of death are of interest for the WHO,

---

<sup>1</sup> December, 26<sup>th</sup>, 2011.

as they are considered one of the main information sources to highlight the diseases and injuries of a given population, in a given area, and for a defined time interval. With an analysis over such data, it enables production of health-related statistics and indicators. In Brazil, this data is stored in the national Brazilian Mortality System (SIM) [5], and grouped by the main causes of death. Both databases are used to monitor health related events, like injuries or diseases, e.g. dengue fever, chagas disease, lymphatic filariasis, among others. [6].

If the goal is to leverage on synergies resulting from querying and comparing the two databases at the same time, ontologies can play an important role in defining a common communication channel needed to ensure semantic interoperability. As mortality and morbidity databases eventually lack uniformity in schema and data levels, ontologies can be used for aggregating common characteristics and, as a source of domain knowledge, for decision support and reasoning [7]. Therefore, the purpose of the current study is to ontologically formalize foundational disease processes and other lifecycle related processes, in the scope of the *Neglected Tropical Disease Ontology* (NTDO) [8] [9].

Ontologies, from a formal point of view, intend to describe a consensus on the nature of entities in a given scientific domain, independently of linguistic variation of the terms used in human communication. Accordingly, formal ontologies are expressed by means of a formal semantics, like Description Logics (DL) [1], nowadays generally using the World Wide Web Consortium (W3C) recommended exchange syntax Web Ontology Language (OWL) [10].

Integrating heterogeneous databases, such as SIM and SINAN, we are facing many interesting problems. For instance, we observed that the identifiers in both databases do not follow strict rules so as to prevent misidentification and to leverage data integration. This syntactical problem is usually addressed by algorithms that compute cumulative evidence [11], from other pieces of the registers to decide for a matching, i.e. comparing other data then the proper identification of the individual (e.g. mother's name, birth date, among others).

However, a more interesting semantic integration problem occurred while querying the two databases together: an individual may happen to die due to a certain disease, but instead of reporting that particular disease as the main cause of death, a typical symptom of the disease was reported as the cause and carried into the database.

For instance, a person who was long-term infected by a *Tripanosoma cruzi* and developed Chagas Disease, suddenly died of a "heart attack". A heart complication, in a bearer of Chagas Disease, is an indirect effect of the infection. Unfortunately, sometimes deaths by heart attack are registered mistakenly as the main cause of death in bearers of Chagas Disease, in the case the disease causes the heart dysfunction.

Following this, a new requirement of the NTDO consists in encompassing the profile of an individual, e.g. human, from life to death in case he or she had been affected by certain diseases, in our case the so-called Neglected Tropical Diseases (e.g. Chagas Disease, Leishmaniasis, Yellow Fever, Dengue Fever, Lymphatic Filariasis, among others). 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 (i.e. as life itself is not characterized easily), the ontological problem of preserving identity of an individual when transitioning from a person to a dead body to dust, among others.

In order to support the integration and verification of morbidity and mortality data in the SINAN and SIM databases, we here present an ontological computer-interpretable representation of death. Also, and as an example of iterative modeling, we outline four versions for representing mortality and discuss representational problems or the complexity of reasoning arising from each. We conclude the article by briefly describing our use case on the integration of the morbidity and mortality databases described above.

## **Methods**

### *Representational Principles*

NTDO [9] leverages on classes and relations provided by the upper level ontology BioTop [12], specializing it downwards to the required leaf node granularity. Additional classes and relations mainly for representing time intervals and their boundaries were imported from the General Formal Ontology (GFO) [13, 14].

NTDO was based in established ontology construction guidelines [15], which suggested the untangling of asserted graphs into disjoint orthogonal axes, letting a DL reasoner maintain the tangled poly-hierarchy. Naming conventions provided by [16] were applied consistently.

### *Representation Language and Semantics*

NTDO has been built using the Ontology Web Language 2 (OWL2), which is the World Wide Web Consortium (W3C) [10] recommended syntax, extended with the agreement operator ( $\doteq$ ) [1]. The semantics of OWL 2 is based on Description Logics (DL) semantics [1], and the for an interpretation  $I$  is as follows. If  $\mathbf{f}$  and  $\mathbf{g}$  are properties or role chains, an OWL 2 interpretation  $I$  extended with the following interpretation function:

$$(\mathbf{f} \doteq \mathbf{g})^I = \{a \in \Delta^I \mid \exists b. (a, b) \in \mathbf{f}^I \wedge (a, b) \in \mathbf{g}^I\}$$

where  $\Delta^I$  is the domain of the interpretation.

The mortality representation was edited via the ontology editor Protégé v.4.1 using the embedded reasoner HermiT [17] for auto-classification. Inference could not be performed over agreements since Hermit is not able to deal with it.

### *Knowledge Sources*

As for the knowledge sources, apart from the literature review, other relevant sources were the Morbidity and Mortality systems themselves [4, 5]. At some extent we grounded our definitions on the way death cases are reported to the SIM [5]. This system reports cases objectively by “main cause of death” and “other related causes of death”, always identified by a forensic medicine service or the physician who was treating the patient for a disease, or injury, leading to the death.

As morbidity and mortality databases do contain homologous entries referring to the same person, in cases where the cause of death of a given individual was previously identified, both

entries shall be semantically related so that a causal profile can be drawn, e.g. postulating the previously registered disease taken from one database as the cause for the death of the same patient in the other database. Such causal connections are explicitly represented in the ontology, without binding to any process of notification via a system. It was done in order to comply with the principles of minimal encoding bias and minimal ontological commitment [18]. This design decision was taken for fostering its reuse by other applications.



## Results

In this section, we describe the ontological representation of mortality. This model assumes a disease to be the primary cause of death and is necessary to describe the processes and events (which are ultimately instantaneous processes) starting with the transmission of a pathogen, over the disease as a pathological process, and finally ending in the process of dying. Before continuing, we must explain informally the distinction between processes and events: the former occurs in a time interval (e.g. a disease) while the latter takes place in an instant. Indeed, an instantaneous process is a process which occurs in a point of time rather in an interval. Axioms illustrating this difference are provided in the next section (version 2).

In the next subsections, we will provide DL definitions for all important parts of our Mortality model, and represent complex issues encountered and how they were solved in our model.

### *Representing Injury and Death*

The representation of death which will be presented in this work is based on the lifecycle displayed in Figure 1 below.

Taking the birth as the starting point, the lifespan of an individual organism continues until its death. However the lifespan may overlap at the end with the beginning of biological death processes, one or more of which will ultimately lead to its death. This definition is grounded on the fact that several factors can simultaneously influence the lifetime of an individual organisms and provoke its death, such as an accident or an illness, or both e.g. a stroke caused deadly traffic

accident. Despite being a simplification only introduced for the easy handle of statistics, many mortality registries worldwide are able to capture and store only one single main cause of death. In our ontology we follow this simplification, although extending it to accommodate multiple causes would be a quite straightforward process without additional computational costs for querying.

At a given moment an individual organism can acquire a certain disease, *e.g.* dengue fever, which may cause premature death, depending on the disease, comorbidities and the health conditions. The endpoint of existence of an individual is usually attached to one or more causes, which are determined by medical investigation. In medical terms, a disease cause is a function of the physiological state of the individual; it can be related to previous biological death processes, registered or not, in the life of an individual.

Assuming all data is available, it should be possible to describe and trace the sequence of causally induced and at times overlapping pathologic processes which affect the life of that organism, from birth to death. Some of them may damage the organism's overall physiological state to such an extent, that they directly initiate a process of physiological death, leading to death itself. This sequence of processes is sometimes evidenced by the records of an individual when the cause of death was previously registered in a morbidity system, *i.e.* the primary cause of death was already known.

We base our (necessary but not necessary and sufficient) definition of 'life' on the work by Koshland [19], who identifies its key characteristic to be

- a) the inherent capability of development according to a program(as described in the DNA) in normal conditions of life.
- b) the accomplishment of kinetic interactions (such as genome and metabolism)
- c) the capability to mutate and enable selection via survival of the fittest individuals.
- d) 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 segregation.

Our definition of the 'Birth' process is based on the description of "live birth" provided by the Brazilian Institute of Geography and Statistics (IBGE) [20]. 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, breathes or exhibits some other vital signs, *e.g.* heartbeat, voluntary muscle contraction, umbilical cord contraction, regardless of the cord being cut or not, and whether or not the placenta was expelled. Conversely, "death" as a state means absence of brain functions and cessation of all biological functions, inherent to the human body [21].

However, there are major difficulties related to the accurate representation of the processes that make an individual die:

- Complexity is an issue, as the causal nature, which can be quite indirect at times with many unknown factors as comorbidities and interlaced parallel influences converging ultimately into a death process.
- Another issue is relating sequences of processes and time, with a precise description of when each process took place, when it started and where its borders are.

Nevertheless, this exact information is probably not important at all if the aim of the model is to deal with mortality data. Instead what is usually known and found in the databases is the knowledge of what is the sequence of typical signs and symptoms of a disease, because the time constraints involving them, e.g. during tuberculosis, a cough with secretion is followed by a pulmonary infection, can be checked in morbidity and mortality notifications. For stating a death record in a mortality notification database, viz. SIM, a physician certifies the underlying cause of death and sometimes secondary ones. The ontology should support these two descriptions.

We also assume the notion of instantaneous processes available in BioTop as equivalent to 'event' provided in GFO, which makes the patient exhibit a certain behavior which is linked, causally or not, to some processes [14].

Next, we present the main challenges related to the representation of domain entities and the logical axioms characterizing and solving these challenges. For this purpose, the representation is divided into four versions, which demonstrate our iterative optimization approach and the evolution of the model to a final proposal.

### ***Representational Challenges of the Mortality Model***

Several challenges were encountered in creating a coherent representation for a mortality process, such as preserving 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 user queries. Each of these items is discussed in the consecutive versions until we arrive at a satisfactory model.

## Version 1

Our initial naïve definition of death was:

*DeathProcess* equivalentTo *InstantaneousProcess*

and (**hasLocus** some *GeographicLocation*)

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

and (**processualPartOf** some *BiologicalDeathProcess*)

and (**hasInstant** some *PointInTime*)

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

This definition lacks precision regarding how to preserve identity between the living and the dead organism, as the living individual is not specified in the axiomatic description. According to the class definition, there is no guarantee that the living and the dead body are identical, since the patients of the *DeathProcess* and *BiologicalDeathProcess* may not be the same.

Also, the axiom expresses no cardinality constraint, which gives rise to different interpretations, such as the possibility of more than one individual dying by the same death process.

Besides, subscribing to the idea that a living organism is eventually transformed into a dead one causes further representational problems. First, our imported top level, BioTop, restricts its organism hierarchy to living ones, requiring additional class expressions to refer to dead organisms (e.g. using the relation **transformationOf**).

As a consequence, a dead human is not 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, i.e., once an individual is an instance of a rigid class, then it ceases to be an instance only when it does not exist anymore [22]. Even if we assume that this description corresponds to a *phased sortal* [22], i.e. entities which change phase (from “living” to “dead”), it is not clear until when identity should be preserved, i.e. whether the ashes of a dead organism should be identified with the dead person both of which could be named Hans ?

## Version 2

A solution to circumvent such representational problems is simply not separately representing the entities that cause this confusion, viz. *DeadOrganism*, which, indeed, do not matter in most health-related applications. The new solution then consists in representing living organisms based on their temporal existence, limited by two time points, as described in the General Formal Ontology (GFO) [14]. Such representation employs the definition of *gfo:Chronoid*, i.e. an interval not defined as a set of points, thus implying in time represented as a continuum, which is equivalent to *biotop:TimeInterval*

*biotop:TimeInterval* equivalentTo *gfo:Chronoid* (2)

Every *Chronoid* has two outer boundaries, known as time limits (gfo: *TimeBoundary* or biotop: *PointInTime*) or points in time. In GFO, there are two kinds of temporal boundaries, representing the right and left limit of a temporal interval, i.e. gfo: *LeftTimeBoundary* and gfo: *RightTimeBoundary*. By definition, they cannot hold the same values in a single chronoid [14]. A schematic representation can be found figure 2.

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

*gfo:Chronoid* subclassOf

(*gfo:hasLeftTimeBoundary* exactly 1 *gfo:LeftTimeBoundary*) and (3)

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

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

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

*gfo:LeftTimeBoundary* equivalentTo *gfo:TimeBoundary* (5)

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

When there are chronoids 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 chronoids represents the notion of temporal region [14].

Following the GFO perspective, entities that exist in a time interval are referred to as *gfo:Presential* (Herre et al., 2007). This class includes material objects that are represented in BioTop as *biotop:MaterialObject*. Thus, the former can be regarded as superclass to the latter.

Following this assumption, the axioms below should be included:

*biotop:MaterialObject* subClassOf *gfo:Presential* (5)

*biotop:LivingOrganism* subClassOf (*gfo:exists\_at* exactly 1 *gfo:TimeBoundary*) (6)

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

Additionally, processes are projected (*gfo: projectsTo*) to *Chronoids*, i.e., they exist in the time interval represented by a *Chronoid* [14]. Establishing correspondences between GFO and BioTop to avoid mismatches in NTDO, the class *gfo:Process* must be mapped to the class *biotop:Process*. Table 1, below, briefly describes the mappings created between BioTop and GFO, which are necessary for NTDO.

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



*DeathProcess* equivalentTo *InstantaneousProcess*

and (**hasLocus** some *GeographicLocation*)

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

and (**processualPartOf** some *BiologicalDeathProcess*)

and (**hasnstant** some *PointInTime*)

On the one hand, the ontological problems with the existence of *DeadOrganisms* are solved, including the identity problem, as 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 process of death by injury or disease happen to several people simultaneously.

Moreover, it still contains three identity problems: (a) The one between the *DeathProcess* 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 *DeathProcess* and the end of the *BiologicalDeathProcess* that arrived at it. Indeed, *DeathProcess* is exactly the last part (in temporal terms) of a *BiologicalDeathProcess*; this is also an issue of coherence since the opposite (a *BiologicalDeathProcess* being part of a *DeathProcess*) would mean that an instantaneous process would have as part a process project to a time interval.

### Version 3

The last mentioned problem can only be solved with the representation of instantaneous processes. For this purpose, the class `biotop:InstantaneousProcess` was used, as being a process that happens in the end of a preceding process so as to form a process sequence, connecting the end of one process with the beginning of the next using the DL agreement operator ( $\doteq$ ). 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, while the latter defines a formation rule for a property, which are usually based in property chains [1] as in the case above. In our ontology, we need, for instance, to establish that a certain process ends exactly when another starts; this is denoted by an agreement.

We now need the definition of an instant to ascribe exactly when a death takes place. In order to enable the condition 'instantaneous' (`ntdo: hasInstant`) to be defined as an exact point in time. This can be reached by making instant an event that occurs solely in the right border of its process that it follows, as below:

$$\text{ntdo:hasInstant} = \text{gfo:projectsTo} \circ \text{gfo:hasRightTimeBoundary} \quad (8)$$

It is important to disambiguate *InjuryEvent* and *DeathEvent*. For the description of an injury event, it is necessary to determine where it took place (ntdo:**hasGeographicLocation**), its cause and the injured patient. Injury causes are described here as being caused exclusively by non-biological processes. All of this is ascribed in the axiom below:

$$\begin{aligned}
 &\text{ntdo:InjuryProcess equivalentTo biotop:InstantaneousProcess} \\
 &\quad \text{and (ntdo:hasGeographicLocation some ntdo:GeographicLocation)} \\
 &\quad \text{and (biotop:causedBy some (biotop:ProcessualEntity} \\
 &\quad \quad \text{and (not biotop:BiologicalProcessualEntity)))} \\
 &\quad \text{and (ntdo:hasInjuredPatient some biotop:LivingOrganism)}
 \end{aligned} \tag{9}$$

Despite not being the focus of the current work, which is about deaths caused by diseases, it is necessary to distinguish pathological processes, structures, and dispositions [23]. Disorders are caused by an accident, a lesion, or a fracture and can lead to a disease. Thus, disorders follow injuries.

The new definition of a *DeathProcess* goes below:

$$\begin{aligned}
 &\text{ntdo:DeathProcess equivalentTo biotop:InstantaneousProcess} \\
 &\quad \text{and (biotop:hasLocus some ntdo:GeographicLocation)} \\
 &\quad \text{and (ntdo:hasPatient some biotop:LivingOrganism)} \\
 &\quad \text{and (ntdo:precededBy some biotop:BiologicalDeathProcess)} \\
 &\quad \text{and (ntdo:hasInstant } \doteq \text{ ntdo:precededBy o gfo:hasRightTimeBoundary)}
 \end{aligned} \tag{10}$$

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

and (ntdo:**hasPatient**  $\doteq$  ntdo:**precededBy** o ntdo:**hasPatient**)

It describes where a death event takes place, which deceased organism is its patient, and which process is the primary cause of death. The agreement conditions are the more important ones.

They ensure that the death occurs exactly when the *BiologicalDeathProcess* is finished

(ntdo:**hasInstant**  $\doteq$  ntdo:**precededBy** o gfo:**hasRightTimeBoundary**)) 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 (the last condition).

Finally, completing the ontological representation of mortality, the class

ntdo:*BiologicalDeathProcess* was created to indicate the existence of an aggregate (summation of processes happening in parallel) of not completely known processes that occur to the dying organism, which ultimately corroborate to the death. A biological death process (from disease to death) is a biological processual entity which is caused by an injury (non-biological) or biological process (but, of course, not by biological death processes themselves). It has as patient a convalescent organism and its duration is delimited:

ntdo:*BiologicalDeathProcess* equivalentTo biotop:*BiologicalProcessualEntity*

and (biotop:**causedBy** some (ntdo:*InjuryProcess* or biotop:*PathologicalProcess*))

and (ntdo:**hasPatient** some biotop:*LivingOrganism*) (11)

and (gfo:**projectsTo** exactly 1 gfo:*Chronoid*)

This axiom addresses the processes that occur prior to the death process and after an injury or disease. As for the representation of participants (also described in *DeathProcess*), there is a need to identify the existence of one or more processes, even imperceptible or indirectly related.

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 statement of a physician that took care of the patient until the time of death. In the present ontology, from the axioms so far described, it is possible to establish the following sequence of facts for an organism: illness / injury → biological death process → death.

The axioms formulated up to now mention only causal relationships (*e.g. InjuryProcess* or *DeathProcess*). However, this notion of causality, which is necessary for the representation, is based on the observer of the process, *i.e.* the 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.

In this ontology, this fact is supported by *ntdo:BiologicalDeathProcess* since it allows for the inclusion of more than one cause, and may be extended in *ntdo:DeathProcess*, since we are only taking into account (here) the primary cause, the defining cause of death (which may not be the real one).

The presented model solved the identity problem; nevertheless a hidden problem not related to the representation but to the reasoning still remains: if agreements are not built over property chains of functional properties, then inference becomes undecidable [24].

Another subtle aspect is that biological death processes may occur due to injury and unknown causes, apart from diseases.

#### Version 4

The undecidability problem mentioned just above is 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. In each death event, only one human being is a patient. Unfortunately, BioTop does not define the relation **biotop:hasPatient** as functional. 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 process admits only one instance of patient (*i.e.* person) (Fig. 3). 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 *DeathProcess*, as can be seen below:

$\text{ntdo:DeathProcess}$  equivalentTo  $\text{biotop:InstantaneousProcess}$   
 and ( $\text{biotop:hasLocus}$  some  $\text{ntdo:GeographicLocation}$ )  
 and ( $\text{ntdo:hasDeathPatient}$  exactly 1  $\text{biotop:LivingOrganism}$ )  
 and ( $\text{ntdo:precededBy}$  some  $\text{biotop:BiologicalDeathProcess}$ ) (13)  
 and ( $\text{ntdo:hasInstant} \doteq \text{ntdo:precededBy} \circ \text{gfo:hasRightTimeBoundary}$ )  
 and ( $\text{ntdo:hasDeathPrimaryCause}$  exactly 1  $\text{biotop:ProcessualEntity}$ )  
 and ( $\text{ntdo:hasDeathPatient} \doteq \text{ntdo:precededBy} \circ \text{ntdo:hasConvalescentPatient}$ )

This 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 and relations presented in the axioms for representing death is depicted below (Fig. 4). 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 compare our work with efforts that discuss mortality epistemologically. Although a related work about an ontology of death by Thomasma [25] enlists related terms and provides some connections among them, it does not a sound definition for death.

The vision based on instantaneous processes, which is applied in our NTDO ontology, is also present in . For him, death can only possibly be identified by another person(s). Another argument supporting our point of view, was described by Schrader [26]. His work indicates that Immanuel Kant reported the fact that abstract, hard to define entities (like 'death') can be distinguished by the particular and unique situation against 'space' and 'time'. Such standpoint is fulfilled by the model shown here.

## Conclusion

In the current work, we represented complex processes, characterized by temporal marks, causality, and an objective and explicit representation of entities and processes related to of death. Several representational difficulties were faced, mainly regarding to the types of entities which are difficult to be precisely described by formal ontologies, *viz.* the relationships among death, diseases and injuries.

Our iterative models, exemplified here by four versions of the ontology, aims at stressing the typical problems – such as preserving identity, asserting correct cardinalities and agreements among relations - encountered in representing complex biological events and entities in description logic, as well as pointing out typical solutions.

The NTDO in its current status allows for an accurate description of the processes related to diseases and injuries, including their evolution that ultimately can lead to death. Using it together with other parts of NTDO, as the description of pathogen transmission by arthropod vectors (present in), a complete sequence of processes can be described in detail, starting from the inoculation of a pathogen by a vector, until the death of an individual. Therefore, the ontology,



with the current addition of mortality related contents, may serve many different purposes, such as supporting tutor systems, serving as shared vocabulary in data integration solutions, etc.

Currently, we are elaborating a use case that matches morbidity and mortality databases. The ontology is being used for 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).

## **Competing interests**

We declare to have no 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, neither now or in the future. DS and SS are supported by the Deutsche Forschungsgemeinschaft (DFG) grant JA 1904/2-1, SCHU 2515/1-1 GoodOD (Good Ontology Design) and the Bundesministerium für Bildung und Forschung (BMBF)-IB mobility project BRA 09/006.

None of the authors hold stocks or shares in an organization that may gain or lose financially from the publication of this manuscript, neither now nor in the future.

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

## **Authors' contributions**

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

RF – Development of the ideas concerning to link morbidity and mortality data, also to link mortality and morbidity entities;

DS – Reviewed the manuscript and gave ideas concerning to the content;

SS – Reviewed the manuscript and gave ideas concerning to the content;

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

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

## Acknowledgements

This work was sponsored by the German DFG grant JA 1904/2-1, DFG SCHU 2515/1-1 GoodOD (Good Ontology Design) and German Ministry of Education and Research (BMBF)-IB mobility project BRA 09/006.

## References

1. Baader F, McGuinness DL, Nardi D, Patel-Schneider P: *The Description Logic Handbook: Theory, implementation, and applications*. 1st edition. Cambridge: Cambridge University Press; 2003.
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. The Gene Ontology Consortium: **Gene Ontology : tool for the unification of biology**. *Nature Genetics* 2000, **25**:25-29.
4. **Sistema de Informação de Agravos de Notificação (SINAN)**  
[[http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=21383](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21383)].
5. **Sistema de Informação de Mortalidade (SIM)**  
[[http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=21377](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=21377)].
6. 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.

7. Bodenreider O, Mitchell JA, McCray AT: **Biomedical ontologies**. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. 2005, **78**:76-8.
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. **OWL 2 Web Ontology: Language Document Overview** [<http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>].
11. Camargo KR, Coeli CM: **[Reclink: an application for database linkage implementing the probabilistic record linkage method]**. *Cadernos de saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública* 2000, **16**:439-47.
12. Beisswanger E, Schulz S, Stenzhorn H, Hahn U: **BIOTOP : An Upper Domain Ontology for the Life Sciences**. *Applied Ontology* 2008, **3**:205-212.
13. Heller B, Herre H: **Ontological Categories in GOL**. *Axiomathes* 2004, **14**:57-76.
14. 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.
15. 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 '03*. New York, New York, USA: ACM Press; 2003:121.

16. 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.
17. Motik B, Shearer R, Horrocks I: **Hypertableau Reasoning for Description Logics.** *Journal of Artificial Intelligence Research* 2009, **36**:165-228.
18. Gruber T: **A translation approach to portable ontology specifications.** *Knowledge Acquisition* 1993, **5**:199-220.
19. Koshland DE: **The seven pillars of life.** *Science (New York, N.Y.)* 2002, **295**:2215-6.
20. **Nascido Vivo** [[http://www.ibge.gov.br/home/estatistica/populacao/registrocivil/nascido\\_vivo.shtm](http://www.ibge.gov.br/home/estatistica/populacao/registrocivil/nascido_vivo.shtm)].
21. Miller FG, Truog RD: **Decapitation and the definition of death.** *Journal of medical ethics* 2010, **36**:632-4.
22. 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.
23. 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.
24. Schmidt-Schauss M: *Computational aspects of an order-sorted logic with term declarations.* Lecture no. Berlin; New York: Springer Verlag; 1989:171.
25. Thomasma DC: **The comatose patient, the ontology of death, and the decision to stop treatment.** *Theoretical Medicine* 1984, **5**:181-196.

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

## Figures

**FIGURE 1** – The typical life cycle of a human. This figure includes the main processes and the points they occur.

(PIC 1 File)

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

(PIC 2 File)

**FIGURE 3** – Graphical Model of an *InjuryProcess*, *DeathProcess* and *BiologicalDeathProcess* and relations (**hasInjuredPatient**, **hasDeathPatient** and **hasConvalescentPatient**), according with the processes participant (e.g. *Human*).

(PIC 3 File)

**FIGURE 4** - Connections between NTDO, BioTop and GFO. Many NTDO classes are subclasses of BioTop classes, while some GFO classes were imported from GFO.

(PIC 4 File)

## Tables

**TABLE 1** – Mappings between GFO and BioTop.

Biotop	mapping	GFO
<i>Material Object</i>	<b>subclassOf</b>	<i>Presential</i>
<i>TimeInterval</i>	<b>equivalentTo</b>	<i>Chronoid</i>
<i>PointInTime</i>	<b>equivalentTo</b>	<i>TimeBoundary</i>
<i>Process</i>	<b>equivalentTo</b>	<i>Process</i>