Towards morbidity to mortality ontological representation in Description Logics

Filipe Santana1§\*, Roberta Fernandes1\*, Daniel Schober2\*, Stefan Schulz2,4\*, Zulma Medeiros3,5\*, Fred Freitas1\*

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

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

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

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

5Pathology 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

FF: fred@cin.ufpe.br

# Abstract

**Background**. Although many efforts are being devoted to the development of biomedical ontologies, very few, if any, encompass sound definitions of death. This concept and its related relations are bound to be used in many applications, e.g. data integration of mortality notification systems. In the current work, we represent complex processes, characterized by temporal marks, causality, and an objective and explicit representation of entities related to death events.

**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 the iterative definitions with four versions of the ontology, so as to stress the typical problems encountered in representing complex biological events, as well as pointing out possible solutions. We designed the ontology following the principles of minimal encoding bias and minimal ontological commitment, in order to foster the possibilities of reuse by a wide range of applications.

**Conclusion**. The Neglected Tropical Disease Ontology (NTDO) in its current status includes the mortality definitions described here, and allows for an accurate description of all 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, 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 its mortality contents may serve many different purposes, such as supporting tutoring systems, serving as shared vocabulary in data integration solutions, etc. Currently, we are working on a use case that matches morbidity and mortality databases, as checking whether the notified data is correct against the constraints imposed by the complex axioms (such as impossibility of a certain disease to be acquired 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).

# 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/), researchers have been relying on ontologies for connecting the different senses for a term related to certain pieces of data as well to ascribe precisely the perspectives in which each of the data sources involved employ a given term. In other words, the task of intelligent information integration (\*\*\*) has been facilitated by the use of ontologies, since they assure interoperability in a safe way in terms of semantics of their representational units and disambiguation of many senses of the same term employed in the different databases. Indeed, more than 250 ontologies[[1]](#footnote-1) are available in the BioPortal bio-ontology library [1] for representing many different topics of biomedical interest and, also, to provide intelligent data retrieval from such databases using ontologies together with other semantic web technologies and resources.

Although many efforts are devoted to the development of ontologies, mainly following the structure, relations and annotations provided by the Gene Ontology [2] 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” which aims at eliminate Tuberculosis [3].

As the epidemiological surveillance actions aim at identifying and determining the factors which promote a given disease, 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]. Mortality databases are also related to the problem, once the cause of death may be one of special interest for WHO, such as …. In Brazil, these data are stored in a nationwide mortality system, *viz.* the 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. tuberculosis [6].

If the goal is to query the two databases at the same time, ontologies can play an important role in promoting data and information retrieval, data integration, data exchange and semantic interoperability. As mortality and morbidity databases lack uniformity, ontologies can be used for aggregating 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 human lifecycle related events, 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) [10], nowadays generally using the World Wide Web Consortium (W3C) recommended exchange syntax Web Ontology Language (OWL) [11].

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 to leverage data integration. This is a syntactical problem, which is usually addressed by algorithms that compute cumulative evidence from other available registries, in the same data source, to decide for a matching. However, a more interesting semantic integration problem was brought about while querying the two databases together: an individual may happen to die 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.

EXAMPLE

Therefore, 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. 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.

Therefore, the focus of the article is to attempt an ontological representation of death, including its subtleties. As an example of iterative modeling we outline four versions for representing mortality and interleave this presentation with a discussion of the 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. The death processes and related events incremented to the NTDO play a crucial role for the proposed solution.

# Methods

For the representation of the new definitions, NTDO specializes classes and reuses relations provided by the upper level ontology BioTop [12]. Additional classes and relations were imported from the General Formal Ontology (GFO) [13, 14], mainly for representing time intervals and their boundaries.

The current work was based in established ontology construction guidelines [15], which suggested the untangling of asserted graphs into disjoint orthogonal axes, and the naming conventions provided by [16].

NTDO has been built using the Ontology Web Language 2 (OWL2), as recommended by the World Wide Web Consortium (W3C) [17], based on Description Logics (DL) [10], supported by the ontology editor Protégé v.4.1 and the embedded reasoner HermiT [18] for auto-classification.

As for the knowledge sources, apart from the literature review, other relevant sources were the Morbidity and Mortality systems themselves [4, 5]. The use case addressed is based on an ontological representation of foundational process regarding the human life cycle, from life to death.

At some extent we grounded our definitions on the way death cases are reported to the Mortality Notification System [5]. This system reports cases objectively by “the main cause of death” and other secondary causes, always identified by a forensic medicine service or the physician who was treating the patient, e.g. in a hospital.

As morbidity and mortality databases contain entries for the same person, in case the cause of death of a given individual was previously identified, both entries shall be semantically related so that a causal profile can be stated. Those connections are explicitly represented in the ontology, without binding the former to any process of notification via a system, in order to comply with the principles of minimal encoding bias and minimal ontological commitment [19]. 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 is related to diseases as primary causes which lead to the death of an individual. This representation seems to be necessary to describe the processes and events occurring starting from the transmission of a pathogen, then the 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

The representation of death which will be presented in this work is based on the lifecycle displayed in Figure 2 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 widely known fact that several factors can influence the lifetime of an individual organisms and provoke its death, such as an accident or an illness. Many mortality registries worldwide admit only a single main cause of death, as a kind of a simplified system design, which helps build health statistics. In our ontology we follow this design decision; although extending it to accommodate many causes would be a quite straightforward process without additional computational costs for querying.

At a given moment in the life of an individual organism it can acquire a certain disease, *e.g.* dengue fever, which may cause premature death, depending on the disease and the health conditions. The endpoint of existence of an individual is usually attached to one or more causes, which are determined by medical or medical-legal investigation. In medical terms, a 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.

Thus, it is possible to describe and trace the sequence (or overlapping) of the deleterious processes which pervade the life of that organism, from birth to death. Some of them may damage the organism’s overall physiological state and initiate a process of physiological death, 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. Moreover, 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.

Firstly, 'life' is taken in this work as described by Koshland [20], being related to the inherent capability of being naturally programmed (as described in the 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 segregation.

The definition of 'Birth' assumed by us for this work is based on the description of "live birth” provided by the Brazilian Institute of Geography and Statistics (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 if the 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].

However, there are major difficulties related to the accurate representation of the processes that make an individual die, mainly due to the complexity in relating sequences of processes and time, with a precise description of when each process took place. Nevertheless, this exact information is probably not important at all; instead what counts 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.

We also assume the notion of '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 related to them, as introduced until now. For this purpose, the representation is divided into four versions, which demonstrate 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 event, 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

An initial definition of death could be:

*DeathEvent* equivalentTo *Event*

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

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

and (**hasProcessualPart** some *BiologicalDeathProcess*)

and (**hasInjuryInstant** some *PointInTime*)

indicating that a death event is an event 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 and the relation of identity and cardinalities. 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, which gives rise to different interpretations, such as the possibility of more than an individual dying by the same death process. 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.

Besides, subscribing to the idea that a living organism is eventually transformed into a dead one causes further representational problems. First, 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 [23]. Even if we assume that this description corresponds to a *phased sortal* [23], *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 such representational problems, and not surprisingly the common choice of ontologists who created all other biological ontologies found in the literature, 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,* as entities *sui generis*, i.e. not defined as a set of points, thus implying in time represented as a continuum.

Every *Chronoid* has two outer boundaries, known as time limits (gfo: *TimeBoundary*) 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 assume 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 (2)

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

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

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

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

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

Following this assumption, the axioms below should be included:

biotop:*MaterialEntity* subClassOf *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).

Aditionally, processes are projected (GFO: **projectsTo**) to *Chronoids* [14]. Establishing ocrrespondences between GFO and BioTop to avoid mismatches in NTDO, the class gfo:*Process* must be mapped to the class biotop: *ProcessualEntity*. Table 1, below, briefly describes the mappings created between BioTop and GFO, which are necessary for NTDO.

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

*DeathEvent* equivalentTo *Event*

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

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

and (**hasProcessualPart** some *BiologicalDeathProcess*)

and (**hasInjuryInstant** some *PointInTime*)

On the one hand, the ontological problems with the existence of *DeadOrganisms* are solved, including the identity 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 death by injury or disease happen 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 ntdo:*InstantEvent* was defined, 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 with the beginning of the next using the DL agreement operator (≐). This operator is used in chains of properties to indicate that the instances to be described are connected. In accordance with, we define the following axiom.

ntdo: *InstantEvent* equivalentTo biotop:*ProcessualEntity*

and (gfo:**projectsTo** o gfo:**hasLeftTimeBoundary** ≐

gfo:**projectsTo** o gfo:**hasRightTimeBoundary**) (8)

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

and (ntdo:**hasInstant** some gfo:*TimeBoundary*)

It states that an event will happen in an instant (ntdo:**hasInstant)** and it will always be the junction of the end of a process and the beginning of the same process, since only one instance of the relation gfo:**projectsTo** is allowed.

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 it follows , as below:

ntdo:**hasInstant** = gfo:**projectsTo** o gfo:**hasRightTimeBoundary** (9)

It is worth stressing, the difference between the two operators, ≐ 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 property, which are usually based in property chains [10] as in the case above.

It is important then 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 any other process which affect patients, but not biological 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* (10)

and (not biotop:*BiologicalProcessualEntity*)))

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

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 [24]. 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 *DeathEvent* goes below:

ntdo:*DeathEvent* equivalentTo ntdo:*Event*

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

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

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

and (ntdo:**hasInstant ≐** ntdo:**precededBy** o gfo:**hasRightTimeBoundary**)

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

and (ntdo:**hasPatient** ≐ ntdo:**precededBy** o ntdo:**hasPatient**)

It describes where a death event takes place, which deceased organism is its patient, and which biological 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 ≐** 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 a set 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 or other biological processes (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:*InjuryEvent* or

(biotop:*BiologicalProcessualEntity* (12)

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

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

and (ntdo:**hasInstant** some bfo:*TimeBoundary*)

This axiom addresses the processes that occur prior to the death event and after an injury or disease event. As for the representation of participants (also described in *DeathEvent*), 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.

So far the axioms formulated up to now mention only causal relationships (*e.g.* *InjuryEvent* or *DeathEvent*). 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:*DeathEvent*, since we are only taking into account (here) the primary cause, the defining cause of death (which may not be the real one).

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.

For the purpose of the modeling, our goals seem 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 in role-value-maps (like role compositions or agreements), then inference becomes undecidable [25].

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

### 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**) (13)

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,*as can be seen below:

ntdo:*DeathEvent* equivalentTo ntdo:*Event* (14)

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

and (ntdo:**hasDeathPatient** exactly 1 biotop:*LivingOrganism*)

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

and (ntdo:**hasInstant ≐** ntdo:**precededBy** o gfo:**hasRightTimeBoundary**)

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

and (ntdo:**hasDeathPatient** ≐ ntdo:**precededBy** o ntdo:**hasConvalescentPatient**)

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

# 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 on ontology of death was conducted by Thomasma [26], but without asserting a sound definition for death. His work enlists related terms and provides some connections among them.

The vision based on events is supported by Kment [27]. For him, death can only possibly be identified by external person(s). 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 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 related to events 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.* death and its relationships with injuries, diseases, processes, etc.

The *step-by-step* modeling description displayed here in the four versions of the ontology aim at stressing the typical problems encountered in representing complex biological events, as well as pointing out typical solutions.

The NTDO in its current status allows for an accurate description of all 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 [8]), 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 many purposes, 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). We designed the ontology together with its mortality parts following the principles of minimal encoding bias and minimal ontological commitment [19] in order to foster the possibilities of reuse by a wide range of 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, neither now or in the future. The organizations that few participants is (only DS and SS), are:

* International Bureau Of the German Ministry of Education and Research (BMBF).
* Which DFG project do you mean?

Any 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.

Any of the authors is currently applying for any patents relating to the content of the manuscript.

Any of the authors do not have other financial competing interest.

Any of the authors do not have 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;

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. 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.

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

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

4. **Sistema de Informação de Agravos de Notificação (SINAN)** [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].

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. **THE DESCRIPTION LOGIC HANDBOOK : Theory , implementation , and applications**.

11. **OWL 2 Web Ontology: Language Document Overview**.

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

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

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

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. 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.

24. 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.

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

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

27. Kment a: **On the ontology of biological aging.** *Archives of gerontology and geriatrics* 1988, **7**:221-7.

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

# Figures

**FIGURE 1 -** Connections between NTDO and BioTop and GFO. Many NTDO classes are subclasses of BioTop classes, while some GFO classes were imported from GFO.

(PIC 1 File)

**FIGURE 2** – The life cycle of a human. This figure includes the main processes and the points they occur.

(PIC 2 File)

**FIGURE 3** – The interval described for a gfo:*Chronoid* and its limits (gfo:*LeftTimeBoundary* and gfo:*RightTimeBoundary*), in the temporal axis.

(PIC 3 File)

**FIGURE 4** – Graphical Model of an *InjuryEvent*, *DeathEvent* and *BiologicalDeathProcess*

(PIC 4 File)

# Tables

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

|  |  |  |
| --- | --- | --- |
| Biotop | mapping | GFO |
| *Material Entity* | **subclassOf** | *Presential* |
| *TemporalRegion* | **hasSubclass** | *Chronoid* |
| *TemporalEntity* | **hasSubclass** | *TimeBoundary* |
| *ProcessualEntity* | **equivalentTo** | *Process* |

1. Decenmber, 26th, 2011. [↑](#footnote-ref-1)