

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

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

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

<sup>2</sup>Parasitology Department, Aggeu Magalhães Research Center, Oswaldo Cruz Foundation, (CPqAM/Fiocruz), Recife, Brazil;

<sup>3</sup>Pathology Department, Institute of Biological Sciences, University of Pernambuco, Recife, Brazil;

<sup>4</sup>Institute of Medical Biometry and Medical Informatics (IMBI), University Medical Center, Freiburg, Germany;

\*These authors contributed equally to this work

§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

DS: schober@imbi.uni-freiburg.de

## Abstract

**Background:** Despite the high coverage of biomedical ontologies, very few sound definitions of death can be found. Nevertheless, this concept has its relevance in epidemiology, such as for data integration within mortality notification systems. We here introduce an ontological 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 kinds of processes with blurred or fiat borders that change their type in a continuous rather than discrete mode. Examples of such hard to grasp concepts are life, death and its relationships with injuries and diseases. 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. We point out possible solutions for representing concepts related to biological life cycles, preserving identity of participating individuals, i.e. for a patient in transition from life to death. This solution however required the use of extended description logics not yet supported by tools. We also focus on the interdependencies and need to change further parts if one part is changed.

**Conclusion:** The axiomatic definition of mortality we introduce allows the description of biologic processes related to the transition from healthy to diseased or injured, and up to a final death state. Exploiting such definitions embedded into 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 the identity of the patient.

## Introduction

With the growing need to cope with large-scale biomedical data, researchers have been relying on ontologies to ensure a shared and computer-interpretable meaning of linguistic terms describing such data, fostering intelligent information integration and interoperability [1]. Indeed, more than 250 ontologies<sup>1</sup> are available in the BioPortal ontology library [2].

Despite many efforts devoted to the development of genomics and metabolomics ontologies, often motivated by the prototypical Gene Ontology [3], few are focusing also 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. Morbidity databases, such as the National Morbidity Notification Information System in Brazil (SINAN) [4], are used as the main sources for epidemiological disease surveillance, prevention and control.

Furthermore, mortality databases describing the cause of death are of interest to the World Health Organization (WHO) enabling the production of local or global health-related statistics. In Brazil, this data is stored in the national Brazilian Mortality System (SIM) [5], and grouped by the primary causes of death.

If the goal is to leverage synergies resulting from querying and comparing the two databases at the same time, ontologies can play an important role by enabling a common communication channel needed to ensure semantic interoperability. Rendering the separately maintained

---

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

mortality and morbidity databases accessible via a common ontology allows for synergistic data exploitation, i.e. use of contextual enrichment, consistency checks and reasoning at schema and data level. [6].

The purpose of the current study is to ontologically formalize foundational disease processes and other lifecycle related processes as occurring in the mentioned data sources to expand the *Neglected Tropical Disease Ontology* (NTDO) [7] [8] and ultimately to use NTDO for integrated querying of the Brazilian mortality and morbidity databases.

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) [9].

When trying to integrate heterogeneous databases, such as SIM and SINAN, many interesting problems arose. 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 [10] from other pieces of the registers to decide for a matching, i.e. comparing other data than the proper identification of an 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 the trigger event that ultimately lead to death, i.e. a particular disorder like Chagas disease, a secondary cause, i.e. something else related to the disease like heart attack is reported as the primary cause and stored in the database. This heart complication in a Chagas patient is a frequent secondary effect of the primary cause, the Chagas infection, which should be tracked in the databases like this in order to prevent false epidemiologic measures.

The above requirement leads us to expand NTDO with classes allowing for such granular distinctions and a sound ontological representation of death. Many subtle aspects hamper a precise definition in this case, i.e. a) the conditions in which an individual is considered dead, and b) the ontological problem of preserving identity of an individual when transitioning from a living to a dead organism in different stages of disintegration.

In order to support the integration and verification of morbidity and mortality data in the SINAN and SIM databases, we here present an ontological representation of death. As an example of iterative modeling, we outline four successive versions for representing mortality and discuss representational problems or the complexity of reasoning arising from each. We conclude by briefly describing the what can be done with the mortality representation and the next steps of the NTDO project.

## Methods

### *Representational Principles*

NTDO [8] leverages classes and relations provided by the upper level ontology BioTop [11], specializing it downwards to the required leaf node granularity. Additional classes and relations for representing time intervals and their boundaries were imported from the General Formal Ontology (GFO) [12, 13].

NTDO was based in established ontology construction guidelines [14], which suggested the untangling of asserted graphs into disjoint orthogonal axes, letting a DL reasoner maintain the tangled poly-hierarchy. Naming conventions provided by [15] 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) [9] 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  $I$  is 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.

For instance, if we want to describe all male children named with his father name as first name, the class of people who has the same name as his father can be described as: he is *Human* and **hasFirstName**  $\doteq$  **hasChild**  $\circ$  **hasFirstName**. It means that the first name of the child must be the same as his father's name.

The mortality representation within NTDO was edited via the ontology editor Protégé v.4.1 using the embedded reasoner HermiT [16] for classification in most steps. Inference could not be performed over agreements since OWL2, and so HermiT, is not able to handle 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 “primary cause of death” (e.g. a disease or an injury which triggered a chain of pathological events and led to death) and “other related causes of death” (e.g. other related disease of injury related to the death). The deaths are always identified by a forensic medicine service or the physician who was treating the subject of care for a disease, or injury, leading to the death.

As morbidity and mortality databases do contain homologous entries referring to the same subject of care, the primary cause of death in the mortality database can be correlated with the disease entry of the same subject of care in the morbidity database, postulating a causal relationship; i.e. the tracked primary death cause might be a secondary symptom of the progressing disease.

## Results

In this section, we describe our ontological representation of mortality. It assumes a disease/injury to be the primary cause of death and is necessary to describe both temporarily extended and instantaneous processes in health care relevant life cycle stages, starting with the transmission of a pathogen, over the disease as a pathological process, and finally ending in the process of dying. 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*

Our representation of death is based on the generalized lifecycle displayed in Fig. 1. 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 a biological death process, the one which will ultimately lead to 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 accidents or illnesses.

Despite being a simplification only introduced for the easy handle of statistics, mortality registries worldwide store mainly the primary 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 inferences.



At a given moment an individual organism can acquire a certain disease, *e.g.* dengue fever, which may cause premature death, depending on the circumstances. In medical terms, a disease cause is a function of the physiological state of the individual. There can be a causal link from a disease and its symptoms to a later death process.

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 organisms' 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.

Our definition of the 'Birth' process is based on the description of "live birth" provided by the Brazilian Institute of Geography and Statistics (IBGE) [17]. 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 [18].

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 where and when each process took place, when it started and where its boundaries are.

Nevertheless, this exact information is probably not important at all if the aim of the proposed 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 follows 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 one underlying primary cause of death and sometimes one or more secondary causes. The ontology should support these two descriptions.

We also assume the notion of instantaneous processes available in BioTop as equivalent to events provided in GFO, which makes the subject of care exhibit a certain behavior which is linked, causally or not, to some processes [13].

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

Next, we present the main challenges related to the representation and the logical axioms characterizing and solving these challenges. To allow the reader to follow our lines of reasoning

we explain four successive versions of definitions for the core entities, demonstrating our iterative optimization approach and the evolution of the model to a final proposal.

The two major challenges encountered in creating a coherent representation for a mortality process were the preservation of the identity of related individuals by setting cardinalities, and the rendering of the resulting ontology in a decidable DL. Each of these items is discussed in the consecutive versions until we arrive at a satisfactory model.

### Version 1: Introducing the Death Representation

Our initial naïve definition of death was:

*DeathEvent* equivalentTo *InstantaneousProcess*

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

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

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

indicating that a death event is an instantaneous process (i.e. it happens in the very moment when the person dies) in which a dead organism is a participant. It also states that there are one or more biological processes (e.g. a disease or an injury) as part of the death event and it is not temporarily extended.

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 instances of *DeathEvent* 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 [19]. Even if we assume that this description corresponds to a *phased sortal* [19], i.e. an entity which changes phases (from “living” to “dead”), it is not clear unless identity is be preserved, i.e. whether the ashes of a dead organism should be identified with the dead person.

## **Version 2: Representing Death in the temporal axis**

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) [13]. Such representation employs the definition of *gfo:Chronoid*, i.e. an interval not defined as a set of points, thus implying that time is represented as a continuum, which is equivalent to *biotop:TimeInterval*

$$\text{biotop:TimeInterval equivalentTo gfo:Chronoid} \quad (2)$$

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 [13]. A schematic representation can be found in Fig. 2.

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

$$\begin{aligned} &\text{gfo:Chronoid subclassOf} \\ &\quad (\text{gfo:hasLeftTimeBoundary exactly 1 gfo:LeftTimeBoundary}) \text{ and} \quad (3) \\ &\quad (\text{gfo:hasRightTimeBoundary exactly 1 gfo:RightTimeBoundary}) \end{aligned}$$

$$\begin{aligned} &\text{gfo:LeftTimeBoundary equivalentTo gfo:TimeBoundary} \quad (4) \\ &\quad \text{and (gfo:leftTimeBoundaryOf some gfo:Chronoid)} \end{aligned}$$

$$\text{gfo:LeftTimeBoundary equivalentTo gfo:TimeBoundary} \quad (5)$$

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 one; the overlap representing the beginning of a new chronoid and the end of the previous.

Following the GFO and BioTop perspectives, *LivingOrganism* is represented as a *MaterialObject*. In order to define that a *LivingOrganism* can die, we need to specify that its existence is delimited, which is described in GFO.

Following this assumption, the axiom below should be included:

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

stating that one living organism in exists in only one time interval (its lifespan).

Additionally, processes are projected (gfo: **projectsTo**) to *Chronoids*, i.e., they exist in the time interval represented by a *Chronoid* [13]. Establishing correspondences between GFO and BioTop to avoid mismatches in NTDO, the class gfo:*Process* must be mapped to the class biotop:*Process*. Fig. 4 indicates 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 *InstantaneousProcess*

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

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

and (**hasInstant** 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 a process of death by injury or disease happen to several people simultaneously.

Moreover, it still contains three further identity problems: (a) The one between the *DeathEvent* and the *BiologicalDeathProcess* patients; (b) the set of definitions stated up to that point neither includes the moment of death nor synchronizes it with 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* converging into it. Indeed, *DeathEvent* is exactly the last temporal part of a *BiologicalDeathProcess*; this is also an issue of coherence since the opposite (a *BiologicalDeathProcess* being part of a

*DeathEvent*) would mean that an instantaneous process would have as part a process related to a time interval.

### Version 3: Introducing the agreement operator to enable identity

The representation of instantaneous processes allows us to render the *RightTimeBoundary* of a *BiologicalDeathProcess* synchronous with the *DeathEvent*. For this purpose, the class *biotop:InstantaneousProcess* was used, as being a process that happens at 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 one, 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 is usually based on 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 its cause and the injured subject of care. Injury causes are described here as being caused exclusively by non-biological processes. All of this is ascribed in the axiom below:

$$\begin{aligned}
 &\text{ntdo:InjuryEvent subClassOf biotop:InstantaneousProcess} \\
 &\quad \text{and (biotop:causedBy only (biotop:ProcessualEntity} \\
 &\quad \quad \text{and (not biotop:BiologicalProcessualEntity)))} \\
 &\quad \text{and (ntdo:hasInjuredPatient only biotop:LivingOrganism)}
 \end{aligned} \tag{9}$$

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 [20]. 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:

$$\begin{aligned}
 &\text{ntdo:DeathEvent equivalentTo biotop:InstantaneousProcess} \\
 &\quad \text{and (ntdo:hasPatient some biotop:LivingOrganism)} \\
 &\quad \text{and (ntdo:precededBy some ntdo:BiologicalDeathProcess)} \\
 &\quad \text{and (ntdo:hasInstant } \doteq \text{ ntdo:precededBy } \circ \text{ gfo:hasRightTimeBoundary)} \\
 &\quad \text{and (ntdo:hasPrimaryDeathCause exactly 1 biotop:ProcessualEntity)}
 \end{aligned} \tag{10}$$

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

It describes 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 subject of care (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 in the dying organism, which ultimately trigger the death event. 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 an organism and its duration is delimited:

ntdo:*BiologicalDeathProcess* subClassOf biotop:*BiologicalProcessualEntity*

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

and (ntdo:**hasPatient** only 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 *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 primary cause of death in a *post mortem* analysis (by autopsy, for instance) or the statement of a physician who was taking care until the time of death. In the present ontology, from the axioms so far described, it is possible to assume a causal 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 *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 primary cause of death and sometimes secondary ones.

In this ontology, this fact is supported by `ntdo:BiologicalDeathProcess`, because this class allows for the inclusion of more than one cause, and may be extended in the `ntdo:DeathEvent`, since we are only taking the primary cause into account here, 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 [21].

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

#### Version 4: Ensuring Identity with transitive object properties

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 instance of *LivingOrganism* stands in a **hasPatient** relation to the *BiologicalDeathProcess*.

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 has only one instance of *LivingOrganism* (Fig. 3) as a passive participant. 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 `biotop:InstantaneousProcess`

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

and (`ntdo:hasDeathPatient`  $\doteq$  `ntdo:precededBy` o `ntdo:hasConvalescentPatient`)

and (`ntdo:precededBy` some `ntdo:BiologicalDeathProcess`) (13)

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

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

This definition has the advantage of stressing explicitly the fact that the dead patient and the participant of a *BiologicalDeathProcess*, of two consecutive and linked processes, has the same identity.

For a better understanding, a schematic model highlighting the main classes and relations presented in the axioms for representing a death event is depicted below (Fig. 4). It illustrates the classes and relations needed to model the life cycle of a biotop:*LivingOrganism*, transitioning 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 [22] enlists related terms and provides some connections among them, it does not provide a sound or formal definition for death. The approach to model a death event as subclass of an instantaneous process, which is applied here, is also present in [22]. For him, death can only possibly be identified by another person.

Currently, we are elaborating use cases that match 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 primary causes of death instead of the disease itself).

## **Conclusion**

In the current work, we represented complex processes, characterized by temporal marks, causality, identity preservation of the attending individuals within and the context of an objective and explicit representation of organisms transitioning from life to death. Several representational difficulties were faced, mainly regarding to the complexity of the represented entities.

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

The NTDO in its current status allows the 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, the sequence of processes can be identified, starting from the inoculation of a pathogen by a vector, until the death of an individual.

Unfortunately, the capabilities of reasoners could not be exploited, in the case the representation language they handle and the one chose for NTDO (i.e. OWL2) does not allow the usage of agreements.

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, among others. The usage in data integration solutions seems to be promising, as mortality and morbidity databases contain erroneous and/or incomplete entries.



## **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.

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;

FF – 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;

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

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

## **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. Publication was supported by the German Research Foundation (DFG) within the funding programme Open Access Publishing. We thank Stefan Schulz, Medical University of Graz, for critically reviewing the manuscript.

## References

1. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider P: *The Description Logics Handbook*. 2nd edition. Cambridge: Cambridge University Press; 2007:624p.
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. Bodenreider O, Mitchell JA, McCray AT: **Biomedical ontologies**. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*. 2005, **78**:76-8.
7. 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.
8. **Neglected Tropical Disease Ontology** [<http://www.cin.ufpe.br/~ntdo/>].
9. **OWL 2 Web Ontology: Language Document Overview** [<http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>].
10. 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.
11. Beisswanger E, Schulz S, Stenzhorn H, Hahn U: **BIOTOP : An Upper Domain Ontology for the Life Sciences**. *Applied Ontology* 2008, **3**:205-212.
12. Heller B, Herre H: **Ontological Categories in GOL**. *Axiomathes* 2004, **14**:57-76.
13. 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.
14. 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.

15. 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.
16. Motik B, Shearer R, Horrocks I: **Hypertableau Reasoning for Description Logics.** *Journal of Artificial Intelligence Research* 2009, **36**:165-228.
17. **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)].
18. Miller FG, Truog RD: **Decapitation and the definition of death.** *Journal of medical ethics* 2010, **36**:632-4.
19. 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.
20. 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.
21. Schmidt-Schauss M: *Computational aspects of an order-sorted logic with term declarations.* Lecture no. Berlin; New York: Springer Verlag; 1989:171.
22. Thoma DC: **The comatose patient, the ontology of death, and the decision to stop treatment.** *Theoretical Medicine* 1984, **5**:181-196.

## Figures

**FIGURE 1** – The typical life cycle of an organism. This figure includes the main processes and events from conception to death.

(PIC 1 File)

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

(PIC 2 File)

**FIGURE 3** – Graphical model of an *InjuryEvent*, *DeathEvent* and *BiologicalDeathProcess* and relations (**hasInjuredPatient**, **hasDeathPatient** and **hasConvalescentPatient**), according with the processes, respective death events (InstantiousProcess) 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)