The *Derives_From* Relation in Biomedical Ontologies

Mathias BROCHHAUSEN[a,1]
[a] *Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany*

Abstract. This paper is part of the ongoing efforts within the framework of the Open Biomedical Ontologies (OBO) ontology library to optimize the treatment of relations in biomedical informatics. When ovum and sperm fuse to form a zygote, then the latter is *derived from* the former. This *derives_from* relation is of huge importance not only in embryology but also in every other science dealing with ontogenetic processes. This study examines critically the treatment of *derives_from* in the OBO Relation Ontology in light of its conformity to the underlying processes of development and transformation in biological and medical reality.
Keywords: biomedical ontology, relations, embryology, erythropoiesis

1. Introduction

In recent years the use of ontologies for the development of medical information systems has witnessed an explosive growth [1]. However, initial attempts to use ontologies to support information management in biomedicine have focused primarily on the formulation and definition of *terms*. In ontologies and terminologies such as the Gene Ontology [2], SNOMED [3] the formal treatment of the *relations* which would link these terms has been neglected. [4] puts forward a methodology for the selection and definition of relations in biomedical ontologies, conformity with which has now been adopted as a criterion for inclusion into the OBO Foundry collaborative experiment in ontology development recently initiated by the Open Biomedical Ontologies (OBO) consortium [5]. The benefit of a relation ontology in clinical research has been argued for in [6]. I believe that the approach advanced in [4] is vitally important for the enhancement of biomedical information systems in the future; however, some aspects of its treatment of relations need to be improved if it is not to lead to mistakes when applied to biological or medical cases.

This paper deals with foundational aspects of the Relation Ontology (RO). The aim is to demonstrate that certain formal solutions presented in [4] can lead to confusing results regarding medical practice and biomedical theory. Though computational consequences are not discussed in this publication, it will become obvious that changes in the foundation will lead to changes in application and implementation.

[1] Corresponding Author: Mathias Brochhausen, IFOMIS, Universität des Saarlandes, Postfach 15 11 50 66041 Saarbrücken, Germany, mathias.brochhausen@ifomis.uni-saarland.de

The Relation Ontology distinguishes, in its initial version, ten relations between types of entities in reality, which are labeled as follows:

Foundational relations
is_a
part_of
Spatial relations (connecting one entity to another in terms of relations between the spatial regions they occupy)
located_in
contained_in
adjacent_to
Temporal relations (connecting entities existing at different times)
transformation_of
derives_from
preceded_by
Participation relations (connecting processes to their bearers)
has_participant
has_agent

It is the *derives_from* relation which is the object of the present study. This relation is of special interest given its application in the field of reproductive medicine, where its proper treatment may even bear on issues addressed in biomedical ethics [7]. From an ontological point of view, reproduction and embryological development are interesting since they provide the basis not only for the creation of new organisms but also for the coming into being of new organism parts, as well as for pathological formations such as tumors or blisters [6].

2. Ontological relations in reproduction and erythropoiesis

An informal definition of *derives_from* as conceived by the RO might read as follows:

> One universal C derives from another universal C_1, if the instances of C_1 (c_1) exist prior to the instances of C (c) and the instances C derive from instances of C_1. c derives from c_1 if c and c_1 are not identical, c inherits a biologically significant portion of matter from c_1 and the special region occupied by c in the beginning overlaps with the spatial region occupied by c_1 [4].

Obvious examples of derivation as thus defined are the relations between sperm, ovum, zygote and blastomere. The zygote *derives from* sperm and ovum. A case of derivation in which the derivate derives from two entities is called fusion. Through a mitotic cell division the zygote changes into two blastomeres.

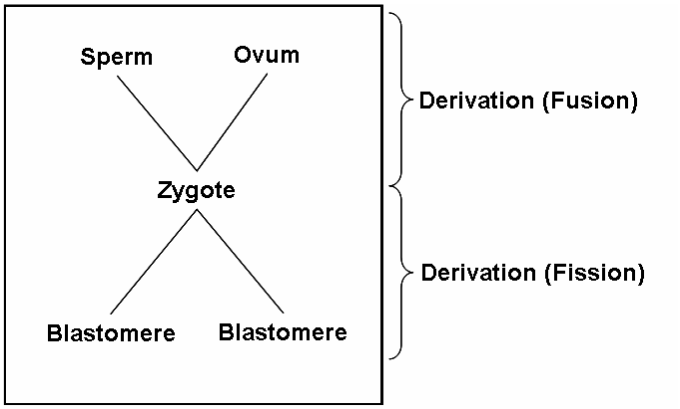

Figure 1: The relation *derives_from* in reproduction (fusion and fission), with time proceeding downwards

Cell divisions then lead progressively to the morula, and this is followed by the formation of the embryoblast, deriving from a number of cells located in the center of the morula. Other cells form the trophoblast. Thus, two fusions are taking place which structure the cells of the morula.

It is crucial to the methodology of the OBO Relation Ontology that a distinction be drawn between relations obtaining between instances, for example the **part_of** relation between John's heart and John's body, and relations obtaining between the corresponding universals or types, for example the *part_of* relation between the types *human heart* and *human body*. The goal of RO is to establish relations between universals. This reflects the basic aim of science, which is to discover general properties or general rules of nature. [4] shows that, from OBO's realist point of view, the only way to formulate relations on the level of universals is to rely on certain underlying relations between the corresponding instances. (This reflects the Aristotelian perspective, according to which knowledge of universals is derived from knowledge of instances.) It is therefore crucially important for the RO to distinguish relations on the instance level from relations on the level of universals.

Derives_from is distinguished from another temporal succession relation in RO, the relation of *transformation*, which can be defined informally as follows:

> One universal C is a transformation of universal C_1 if every instance c of C is an instance of C_1 at an earlier time, but there is not time at which c is an instance of both, C and C_1 [4].

On the level of instances, transformation is just identity: when child John is transformed into adult John, then child and adult are identical: it is *one and the same* person that has undergone a series of changes over time. When John's (healthy) lung is transformed into John's carcinomatous lung, then similarly the healthy lung and the carcinomatous lung are identical: one and the same organ has undergone a series of changes over time.

The RO distinguishes continuants, which endure identically through time, from occurrents (processes) which unfold themselves in their successive phases. We will concentrate in what follows exclusively on relations involving continuants. We will use variables as follows:

C, C_1, … to range over continuant universals
c, c_1, … to range over all continuant instances
t, t_1, … to range over instants of time

On the level of universals , transformation can now be defined as follows [4] (where *Cct* stands for "*c* **instance_of** *C* at *t*"):

> *C transformation_of* C_1 = [definition] *C* and C_1 for all *c*, *t*, if *Cct*, then there is some *t*, such that C_1ct_1, and t_1 **earlier** *t*, and there is no t_2 such that Cct_2 and C_1ct_2 only.

(Here and in what follows we use *italics* for relations between universals, and **bold face** for relations between instances.)

The treatment of derivation, in contrast, is more complex on both the instance and the type level. Provisionally, we can say that derivation occurs where continuant instances succeed each other in time but identity is *not* preserved.

The instance-level **derives_from** is a relation leading to a new entity while instance-level **transformation** is just identity. A good example of the difference between these two relations is provided by the case of erythropoiesis.

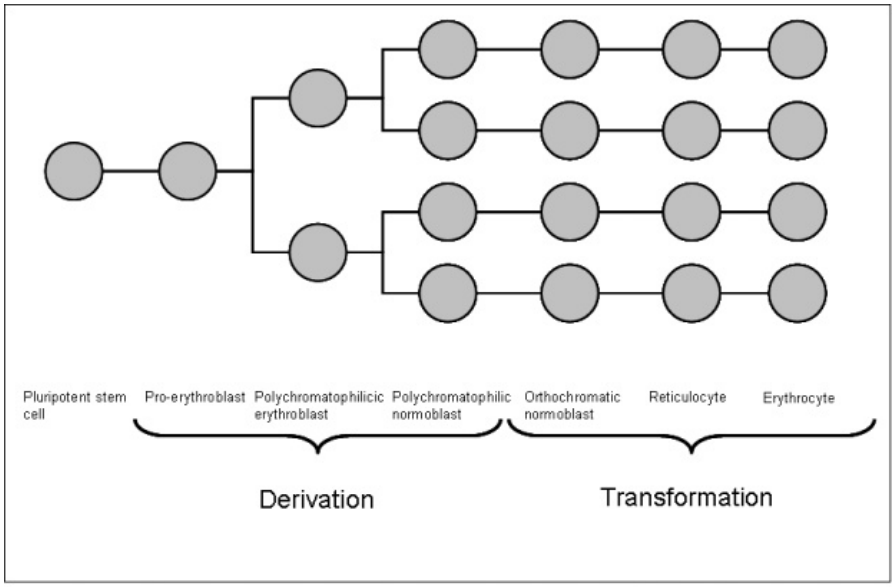

Figure 2: Erythropoiesis [9] This figure shows the development of erythrocytes through mitotic cell division (from pluripotent stem cell to polychromatic erythroblast) and maturation (from polychromatic erythroblast to erythrocyte). (Advance of time is from left to right.)

Erythropoiesis consists of both processes of mitotic cell division and processes of maturation.The first steps in this development consist of **derives_from** relations. Hemocytoblast, proerythroblast, early and late erythroblast all come into existence through mitotic cell division. Thus, fission occurs between these continuants.

The development from late erythroblast to erythrocyte is a process of maturation. Superficially one might think that only **transformation_of** (i.e.. identity) relations

obtain between the continuants, involved. Yet, the ejection of the nucleus marks a **derives_from** relation. The continuants of this relation are the normoblast, the reticulocyte and the pygnotic nucleus.

3. Definition of derives_from

The definition of the derivation relation at the instance level given in [4] is:

> *c* **derives_from** c_1 if *c* is non-identical to c_1, *c* inherits a biologically significant portion of the matter of c_1, *t* such that *c* exists only subsequent to *t* and c_1 exists only prior to *t*.

Three modes of derivation on the instance level are then distinguished: continuation (one entity derives from another entity), fusion (one entity derives from two entities), and fission (two entities deriving from one entity).

The problems with the **derives_from** relation are connected with the inheritance of matter which is referred to in its definition ("inherits a biologically significant portion of the matter of"). For it is not clear what is meant by "a biologically significant portion". A notion which is arbitrary in this way should wherever possible be avoided in a definition.

4. Underdetermination of the *derives_from* relation

A biological example for the problems which result from the existing definition of **derives_from** is the reproduction of Ichneumon flies. Ichneumon lays its eggs in insects, insect larvae or pupae. The hatched larvae of Ichneumon feed on the host, devouring the entire interior and thereby killing the hosts.

The body mass of an ichneumon larva thus derives from the tissue of the host in a way which surely has to be classified as "biological significant". It follows that the whole ichneumon larva derives from the hosting insect larva in the sense of the definition given above. From a genetic point of view, however, the larva derives from its mother, not from the host. Thus, the definition in its present form yields the wrong result.

The larger problems slumbering beneath the superficial observation of the reproduction of ichneumon flies concern the problems which arise when we attempt to deal ontologically with cases of incorporation of biological matter which do not involve incorporation of its genetic code. The example above shows that **derives_from** does not suffice to differentiate between reproduction and ingestion, at least not in all cases known in the biological context.

But more pressing is the question whether **transformation_of** might cover all cases of continuation, which are defined as **derives_from** relations. The example [4] gives for continuation is an incorrect one from the medical point of view. [4] claims that the blastocysts **derive_from** the zygote. Yet, blastocytes do not (directly) derive from zygotes by continuation. The relations involved in the development of the zygote are in fact several fission and fusion types of the **derives_from** relation. The difference between transformation and derivation depends solely on the attribution of identity.

The question has to be raised whether continuation should simply be treated as transformation, especially considering the problems continuation produces in the biological example above. All stages of biological existence from the zygote to the dead organism should be related through **transformation_of** relations and fusion and fission types of the **derives_from** relation.

5. Conclusion

This study shows that the definition of the **derives_from** relation needs refinement. Moreover it has to be doubted whether continuation as one case of **derives_from** should be maintained at all. The curators of RO have agreed to make the necessary changes.

Basically the relations in the biomedical context are **transformation_of** and fusion or fission cases of the **derives_from** relation. The transition from a living person to a corpse is the most referred to example for continuation. But the intuition that there has to be an ontological gap between a living person and a corpse reflects a philosophical or theological attitude, not the attitude of biology and medicine. From the point of view of natural science the relations between a living organism and a dead organism are merely **transformation_of** relations.

Yet, the general outline of the RO ontology should be embraced, since many positive effects have been generated by it. In this paper we have been attempting merely to get some technicalities in a better condition. Improving the ideas developed and promoted by [4] is one way to help formal ontological reasoning to further acceptance and greater success.

References

[1] Rosse, C, Mejino Jr. JLV. A reference ontology for biomedical informatics: The Foundational Model of Anatomy. J Biomed Inform 2003; 36: 478-500.
[2] http://www.geneontology.org/
[3] http://www.snomed.org/
[4] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies. Genome Biology 2005; 6: R46.
[5] http://obo.sourceforge.net/
[6] Smith B, Kumar A, Ceuster W, Rosse C. On carcinomas and other pathological entities. Comp Funct Genom. 2005; 6: forthcoming
[7] Smith B, Brogaard B. Sixteen days. J Med Philos 2003; 28: 45-78.

I am grateful to Barry Smith, Ulf Schwarz and three anonymous reviewers for comments and suggestions for improvement.

This paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the European Union Network of Excellence on Medical Informatics and Semantic Data Mining, and the Volkswagen Foundation under the auspices of the project "Forms of Life".