

Malaria Diagnosis and the Plasmodium Life Cycle: the BFO Perspective

Werner Ceusters and Barry Smith

New York State Center of Excellence in Bioinformatics and Life Sciences,
University at Buffalo, Ontology Research Group, 701 Ellicott street, NY, USA
{ceusters, phismith}@buffalo.edu

Abstract. Definitive diagnosis of malaria requires the demonstration through laboratory tests of the presence within the patient of malaria parasites or their components. Since malaria parasites can be present even in the absence of malaria manifestations, and since symptoms of malaria can be manifested even in the absence of malaria parasites, malaria diagnosis raises important issues for the adequate understanding of disease, etiology and diagnosis. One approach to the resolution of these issues adopts a realist view, according to which the needed clarifications will be derived from a careful representation of the entities on the side of the patient which form the ultimate truthmakers for clinical statements. We here address a challenge to this realist approach relating to the diagnosis of malaria, and show how this challenge can be resolved by appeal to Basic Formal Ontology (BFO) and to the Ontology for General Medical Science (OGMS) constructed in its terms.

Keywords: malaria, Ontology of General Medical Science

1 Introduction

Malaria is a disease caused by one of four types of Plasmodium usually transmitted to humans by the bite of an infected female Anopheles mosquito that previously sucked the blood from a person with malaria.¹ When a patient is or has been in a country in which malaria is prevalent, the disease can be suspected on the basis of both symptoms reported by the patient (such as body aches, headache and general malaise) and physical findings detected at examination (such as severe chills, high fever, and prostration). However, for a definitive diagnosis to be made, laboratory tests must demonstrate the presence within the patient of malaria parasites or their components.²

The parasite life cycle (Figure 1) starts in the human host when Plasmodium sporozoites enter the bloodstream after being transmitted via a mosquito bite. From there, the sporozoites infect the liver cells and disappear from the bloodstream within approximately 30 minutes.

The sporozoites mature into schizonts, which rupture and release merozoites into the blood circulation, where they infect red blood cells. They further undergo asexual multiplication: some merozoites mature again into schizonts that lead to more merozoites; others differentiate into gametocytes which, when picked up by a second

mosquito during a further bite, undergo a series of transformations in this second mosquito, leading eventually to the production of new sporozoites that can infect another human being. Two types of Plasmodium – P. ovale and P. vivax – can persist in the liver of an infected patient and cause relapses by invading the bloodstream weeks, or even years, later.

In some regions people are infected but not made ill by the parasites. This can be so even if the patient manifests symptoms of malaria, which are after all quite unspecific. According to the Centers for Disease Control and Prevention (CDC) ‘*such carriers have developed just enough immunity to protect them from malarial illness but not from malarial infection. In that situation, finding malaria parasites in an ill person does not necessarily mean that the illness is caused by the parasites*’² (emphasis added).

Finally, it is known that patients with sickle cell trait, i.e. people who inherited one sickle cell gene and one normal gene, have a reduced likelihood of dying from malaria because the replication cycle of the parasite is hampered by this condition: infected red blood cells become sickle-shaped and are for this reason destroyed in the spleen along with the associated parasite.

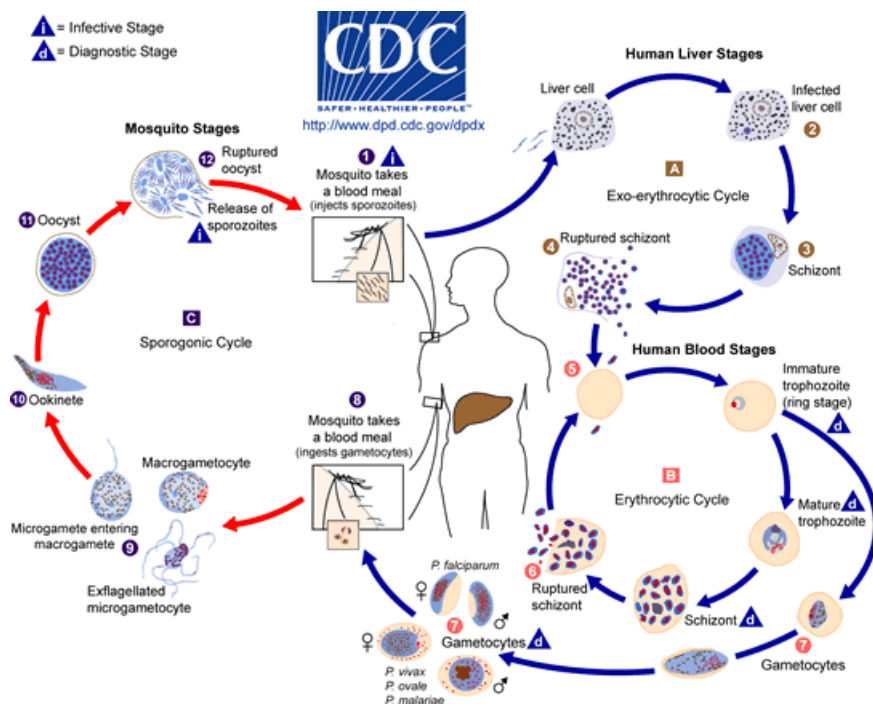


Figure 1: Life cycle of Plasmodium parasites²

2 Background

Basic Formal Ontology (BFO)³ is an *upper ontology* that is intended to provide a logically well-structured set of highly general representational units for common use across multiple scientific and clinical specialisms. BFO is the foundation for the OBO Foundry ontologies⁴ and for a large variety of other *domain ontologies*, especially within the biomedical sphere (<http://www.ifomis.org/bfo/users>).

BFO is designed to serve semantic interoperability of multiple data resources. It is built on a realist basis, which means that it is intended to represent exclusively types of entities that exist in reality, including information entities such as databases or clinical charts, as well as disease entities such as *malaria* or *influenza*. BFO's approach thus differs from those approaches which rest on information modeling and which are said to be 'concept based' or 'object oriented'. Because of important distinctions between the realist and concept-based paradigms, including differences in terminology, communication between the groups on either side is not easy. Thus it has been stated that '*BFO has shortcomings for representing medical information at the granular level. Like many philosophy based upper ontologies, it suffers from defining accidental properties when they are not. This leads to issues in maturation of organisms through development cycles such as parasites go through and leads to erroneous classifications*'.⁵

In response, it must be pointed out that BFO does not use the term 'property' since this term (like its sister terms 'class' and 'concept') is subject to too large a variety of competing interpretations. This does not however mean that one cannot give a definition using the resources of BFO for what is meant by 'property' (or by 'accidental property') as these terms are used in specific contexts. It also does not mean that BFO is for some reason unable to do justice to scenarios involving complex temporal relations between multiple disease-causing processes of the sort exemplified, for example, in the case of malaria.

To test the validity of this latter claim, a challenge was proposed to determine whether BFO is capable of representing a scenario under which John, a person on rotation living in the Congo, has a blood test drawn that shows sporozoites on the smear, in such a way as to be able to answer the following questions:

1. Does John have malaria when there are sporozoites detected on his blood smear?⁵
2. How can BFO be used to classify an immature life form as a cause of a disease when the causative agent develops internally to the organism and changes its stage of life?

3 Objectives

Our purpose in addressing this challenge is (a) to document the sorts of misunderstandings that advocates of the concept orientation have manifested in their approaches to BFO, (b) to highlight the kinds of questions that BFO-users have to ask themselves when analyzing real-life problems, and (c) to demonstrate the benefits of

the realist approach as a robust means of providing upper-level categories in whose terms diverse representations of complex clinical scenarios can be analyzed and compared.

4 Methods

We based our analysis on definitions (Table 1) from the on-line version of the Stedman Medical Dictionary¹ and from the web pages on malaria maintained by the CDC,² including <http://www.cdc.gov/Malaria>. We used the BFO-based Ontology of General Medical Science⁶ (Table 2), formal relationships defined in the Relation Ontology⁷ and additional upper ontology representational units taken from BFO³. Important here are the distinctions between (1) *universals* and *particulars*, for instance *HUMAN BEING* versus John, (2) *continuants* and *occurrents*, for instance sporozoites versus the transformations they undergo through time, (3) *first-order entities* on the side of the patient and those *information entities* which are *about* such first-order entities, for instance John's disease versus some diagnosis made about that disease, (4) *qualities* and *dispositions*, for instance an organism's temperature or mass versus its potential to undergo certain processes when trigger conditions are satisfied, and (5) *diseases* and those *predispositions to disease* which belong to a wider group of what are known as 'risk factors'. These distinctions imply an analysis according to which the question *Does John have malaria when there are sporozoites detected on his blood smear?* in fact amounts to three questions:

- (1) *What is denoted by the term 'malaria'?*
- (2) *Does John have what is called 'malaria' in the specified scenario?*
- (3) *What is required to allow a correct diagnosis of what is called 'malaria'?*

Table 1: Malaria-related definitions from Stedman and CDC

Malaria (D1 – Stedman): a <i>disease</i> caused by the presence of the sporozoan Plasmodium in the erythrocyte phase [...] characterized by episodic severe chills and high fever, prostration, and occasionally death or immunologically mediated sequelae
Malaria (D2 – CDC): a serious and sometimes fatal <i>disease</i> caused by Plasmodium falciparum, P. vivax, P. ovale, or P. malariae. People who get malaria are typically very sick with high fevers, shaking chills, and flu-like illness.
Disease (D3 – Stedman): an interruption, cessation, or <i>disorder</i> of body function, system, or organ.
Disease (D4 – Stedman): a morbid entity characterized usually by at least two of these criteria: recognized etiologic agent(s), identifiable group of signs and symptoms, or consistent anatomic alterations.
Disorder (D5 – Stedman): a disturbance of function, structure, or both, resulting from a genetic or embryonic failure in development or from exogenous factors such as poison, trauma, or disease.

5 Results

There are multiple definitions for ‘malaria’ and ‘disease’ in Table 1 above, and this multiplicity is compounded further if other traditional terminologies such as SNOMED CT are added into the mix. How, then, can we reconcile the differences which arise when clinical data are collected on the basis of such conflicting definitions, given that the underlying concepts employed are so diverse?

BFO is designed to provide an answer to this question by providing a common basis for analysis that can be accepted by all of the multiple specialist communities involved.

The results of our analysis of the mentioned scenario are presented in Table 3, whose columns contain indices for easy reference in the discussion that follows, a unique identifier (ID) for the particular entities in reality to which reference is made, a description of each such entity in the terms set forth above and of the relations in which it stands to other entities, and also specifications of timeframes during which these relationships hold. We assumed in this table that a complete malaria cycle is realized in John; deviations from this canonical case are discussed below.

A number of simplifications are made for reasons of space. Thus we gloss over the relationships that obtain between groups and members of those groups, for example between individual sporozoites and the groups they form, and also over the distinction between an organism such as John and the whole formed by John together with associated entities – for example sporozoites – in his interior.

Table 2: Disease-related definitions extracted from the BFO-based Ontology of General Medical Science (OGMS)

Disease (D6): a disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism.

Disease Course (D7): the totality of all processes through which a given disease instance is realized.

Disorder (D8): a combination of physical components of or in an organism that is clinically abnormal.

Manifestation of a Disease (D9): a bodily feature of a patient that is (i) a deviation from clinical normality that exists in virtue of the realization of a disease and (ii) is observable.

Pathological Process (D10): a bodily process that is clinically abnormal [which is what OGMS says; and which I prefer; let me know if this causes you problems].

Predisposition to Disease of Type X (D11): a disposition in an organism that constitutes an increased risk of the organism’s subsequently developing disease X.

Diagnosis (D12): a conclusion of an interpretive process that has as input a clinical picture of a given patient and as output an assertion to the effect that the patient has a disease of such and such a type.

Table 3: Simplified representation of John’s disease history from infection until his first malarial attack

Ref	ID	Description	Relations	Time
1	#1	John	inst-of <i>HUMAN BEING</i>	t1-t7
2	#2	John’s history	inst-of <i>OCCURRENT</i>	
3			has-participant #1	t1-t7
4	#3	sporozoites in John’s blood	inst-of <i>BONA FIDE GROUP</i>	t2-t3
5			part-of #1	t2-t3
6			inst-of <i>DISORDER</i>	t2-t3
7	#4	disposition to harm John’s liver cells	inst-of <i>DISPOSITION</i>	t2-t3
8			has-bearer #3	t2
9	#5	portion of John’s blood on a blood smear	part-of #1	t2-t3
10	#6	sporozoites in John’s blood smear	inst-of <i>BONA FIDE GROUP</i>	t4-t7
11			part-of #5	t4-t7
12			part-of #3	t2-t3
13	#7	schizonts in John’s liver	part-of #1	t5
14			derives-from #3	t5
15	#8	process in which #7 participates	inst-of <i>PATHOLOGICAL PROCESS</i>	
16			realization-of #4	t5
17			has-participant #7	t5
18	#9	disposition to harm John’s blood cells	inst-of <i>DISPOSITION</i>	t5-t7
19			has-bearer #7	t5
20	#10	mature Plasmodia life forms in John’s blood	part-of #1	t6-t7
21			derives-from #7	t6
22			participates-in #2	t6-t7
23	#11	disposition to cause pathological processes characteristic of malaria	inst-of <i>DISEASE</i>	t6-t7
24			has-bearer #10	t6
25	#12	clinical manifestations of which John is the bearer	part-of #2	t7
26	#13	clinical manifestations of John’s malaria	part-of #12	
27			realization-of #11	t7

6 Discussion

The purpose of the analysis presented here is to demonstrate the power of the realist approach in representing reality, including those parts of reality that can not, for whatever reason, be directly observed. The issued challenge involves a concrete example of entities of this sort. Sporozoites in the blood are very shortlived; they disappear within 30 minutes. Also, due to the very small numbers involved – malaria can be caused even by one sporozoite – sporozoites are almost impossible to detect in the blood. They therefore play no de facto role in the diagnostic process (and diagnosis) of malaria because this diagnosis is only considered after the first or second bout of fever, which occurs days after the infectious mosquito bite and thus certainly not within 30 minutes. One of the elements crucial to the process of accepting malaria as a diagnosis for the disease underlying the manifestations is

indeed the recurrence of fever attacks the first of which initiates the phase of clinically observable manifestations of the disease. It is these recurring fevers that routinely lead doctors to either diagnose a condition as malaria or initiate the relevant diagnostic procedures. Thus, although the question ‘*Does John have malaria when there are sporozoites detected on his blood smear?*’ does not make much sense in light of currently available diagnostic techniques, it still poses interesting problems from an ontological perspective. And for sure, the related question ‘*Does John have malaria when there are sporozoites in his blood?*’ is still an important question to address, since it helps us to guard against a widespread assumption according to which *presence of disease* and *diagnosis of disease* can be taken to be the same thing. That they are different is demonstrated most clearly by the common occurrence of patient records containing multiple diagnoses for what, in course of time, proves to be a single disease. To the best of our knowledge, however, there is to date no electronic health record system that allows the structured documentation not only of diagnoses, but also of what, on the side of the patient, these diagnoses are about.

Following OGMS, to assert that something is *clinically abnormal* is to state (1) that it is not part of the life plan for an organism of the relevant type (unlike, say, aging or pregnancy), (2) that it is causally linked to an elevated risk of pain or other feelings of illness or of death or dysfunction on the part of the organism, and (3) that it is such that this elevated risk exceeds a certain threshold level.^{6,8}

Taken together with **D8** in Table 2, this implies that the sporozoites found in John’s blood smear (#5) provide partial evidence for the presence of a disorder in John. First, physical components of this sort satisfy two of the three criteria for clinical abnormality: presence of sporozoites does not belong to the life plan of a human being, and it is causally linked to an elevated risk of death and dysfunction. However, on the basis of the information provided, no inference can be drawn as to whether the third criterion is also satisfied: for if, *per accidens*, all sporozoites that were in John’s body prior to the blood tap end up on the smear, then there is no longer any risk related to these sporozoites. In that case, there would be no entities to which #7 – #11 and #13 in Table 3 would correspond. If, however, there are further sporozoites in that patient’s body, then John has what OGMS refers to as *a disorder*.

The second question is whether John, at that point, has a *disease* as perceived under the OGMS framework, and if so what entity that disease would *be*. The determining factor, following **D6**, is whether or not the given sporozoites dispose or predispose to pathological processes (**D10**), i.e. processes that either are changes in the way a normal physiological function is realized (e.g. hyperventilation) or have no physiological counterpart at all (e.g. inflammation). Clearly, this question, too, can be answered positively: the sporozoites penetrate liver cells as a result of which the generation of merozoites then starts: these processes have no human physiological counterpart and are thus pathological. The disorder in question is thus the bearer of a *disposition to pathological processes*, which means, in OGMS terms, that it is the bearer of a *disease*.

The third question is whether the disease in question is what is standardly called ‘*malaria*’. Here the answer is less straightforward, and this for a number of reasons. The first is that authoritative medical sources are vague about what it is for something to *be* a disease of this or that sort, as contrasted with what it is for a person to *have* a disease. Often these sources characterize a disease as being an *illness*, *sickness*,

pathological condition, morbid entity, and so forth, whereby these terms themselves are either not further defined, or, if they are defined, then in a way which leads to circularity or inconsistencies when the provided definitions are combined. We could not find a definition for disease to which the CDC adheres consistently in its treatment of malaria, and the definitions in Stedman (**D3** – **D5**) suffer from analogous shortcomings.

On the proposed OGMS definition, a disease is a disposition to pathological processes inhering in a disorder of some specific sort – for instance involving a necrotic liver or a chromosome with abnormal mHTT. The disorder provides the physical basis of the disease and it also provides the *differentia* by which diseases of different types are, from the realist perspective, distinguished from each other. This realist treatment of disease regards manifest symptoms and pathological processes as being, in a sense, epiphenomena. The disease (i.e. the disposition) can still exist even if, for whatever reason, it is not realized in any overt way. Clearly, however, appeals to manifest symptoms and pathological processes will still be needed for diagnostic (epistemological) purposes, since for many diseases we still have very little knowledge of the nature of the disorder which forms their physical basis.

Providing an answer to our third question, now, is less than straightforward against this background because there are, in the described scenario, three distinct dispositions that may qualify as diseases: #4, the disposition to harm John's liver cells, #9, the disposition to harm John's blood cells, and #11, the disposition in John to develop clinical manifestations caused by the presence of mature Plasmodia in his blood cells. The CDC states that '*the blood stage parasites are responsible for the clinical manifestations of the disease [of malaria]*', these manifestations being, for instance, elevated temperature, perspiration, weakness, enlargement of the liver, increased respiratory rate, and so forth. This, together with **D2**, leads us to conclude that it is #11 that would be qualified by the CDC as John's malaria, though there is some evidence that #13 might also be so identified – though this, from the OGMS perspective, would amount to a confusion of a *disease* with the corresponding *disease course* or *series of symptoms*. **D1** and **D4** together lead us to conclude that for Stedman, too, either #11 or #13 would qualify as the disease called 'malaria'. However, on combining **D3** with **D5**, we see that #4, #8, and #9 would also qualify as 'diseases' for Stedman, although not as 'malaria' (because of **D1**'s explicit reference to the erythrocyte phase).

Similar analyses can now be provided for the non-canonical scenarios regarding the unfolding of the parasite's life cycle. For those malarial forms in which parasites survive in the liver, the relationships 13, 14, 16, 17 and 19 extend beyond the period labeled 't5'. For patients with sickle-cell trait or with acquired immunity for malaria the analysis is as follows. Such persons may be infected yet not develop clinical manifestations related to the infection. Were this to apply to John, this would mean that there would be no counterparts of the entities labeled #12 and #13 in Table 3 except in those circumstances where John had another disease whose symptoms mimic those of malaria, in which case a counterpart of #12 would exist. Only if John should exhibit a total and irreversible immunity against the morbid effects of the parasite – he is an *absolute immune carrier* – would disposition #11 also not be present. In case of *partial immunity* carriers are able to maintain a state of homeostasis: pathological processes that are the realization of disposition #11 still

come into existence, but the body is able to prevent their clinical manifestation. In this case, the CDC seems to entertain the proposition that such patients do not have a disease. On the OGMS view, in contrast, John would have the disease, though in a dormant form.

7 Conclusion

Have we been able to successfully respond to the challenge? The answer seems to be: yes. First, while we follow OGMS in the above in defining ‘disease’ in terms of **D6**, we note that an alternative analysis could be formulated in BFO terms on the basis of a definition of ‘disease’ close to that of **D7**. The latter has the advantage that it is favored by many clinicians, but we believe that it faces serious problems in doing justice to the wide variations in clinical presentations which are observed for many diseases.⁶ We believe that it faces problems also in doing justice to the interactions between disease and treatment. On the OGMS view, a disease that is being successfully treated with symptom-suppressant drugs remains one and the same disease, though the disease course is here radically modified.

If the disease of malaria is, as according to **D6**, a disposition, then we need to determine which of the three dispositions #4, #9 or #11 – or perhaps which combination of these three dispositions – is most appropriately called ‘malaria’. If, as according to **D1** and the CDC, #11 is chosen, then #4 and #9 come to be recognized as *predispositions* to malaria, following **D11**. It then still remains open whether they themselves are properly to be classified also as diseases in their own right.

Whichever decision is taken, it is in each case possible to determine whether or not John has malaria under the chosen definition, and if so, *when* he has the disease. We demonstrate this in Table 4 by inspecting three cases: a standard case, a case in which a person is a carrier but has developed total immunity, and a case in which this immunity is only partial.

The question whether, under the distinguished scenarios and definitions, a correct diagnosis can be made depends of course on what at any given time is known about John. If for instance **D6** as exemplified by #9 is chosen as definition for ‘malaria’, and if it is not known that John is totally immune, then the blood smear evidence will likely lead to an incorrect diagnosis to the effect that John has the disease. The additional challenge, i.e. to ‘*classify an immature life form as a cause of a disease when the causative agent develops internally to the organism and changes its stage of life*’, is hereby also met. The causal relationship between the immature life form – the sporozoites – and the occurrence of clinical manifestations is clearly demonstrated by at least the following chain of relationships:

- #1 **has-part** #3 at t2-t3 (rel 5)
- #3 **derivational-source-of** #7 at t2 (rel 14)
- #7 **derivational-source-of** #10 at t6 (rel 21)
- #10 **bearer-of** #11 at t6 (rel 24)
- #11 **realizes** #13 at t7 (rel 27)

Table 4: Specification of the time when John has malaria under three possible definitions and three case types.

Case type	Definition for ' <i>malaria</i> ' based on D6 exemplified by		
	#4	#9	#11
standard	t2-t3	t5-t7	t6-t7
absolute immune carrier	t2-t3	-	t6
partial immunity	t2-t3	t5-t7	t6-t7

It is an open question whether these same challenges can also be met by standard information-modeling approaches to the representation of clinical scenarios illustrated, for example, by the HL7 RIM.

Acknowledgements

Thanks go to Cecil O. Lynch, who issued the challenge, and to Ronald Cornet and Christos (Kitsos) Louis for valuable suggestions. The work described was in part funded by the John R. Oishei Foundation, NIH Roadmap Grant 1U54 HG004028 supporting the National Center for Biomedical Ontology and grant NIH/NIAID R01 AI77706-01 supporting the Infectious Disease Ontology (<http://www.infectiousdiseaseontology.org>).

References

1. Stedman's. Stedman's Online Medical Dictionary. Wolters Kluwer Health; 2009.
2. Centers for Disease Control and Prevention. Diseases & Conditions A-Z Index. National Center for Health Marketing; 2009.
3. IFOMIS. 2009 September 7. Basic Formal Ontology. <<http://www.ifomis.uni-saarland.de/bfo/>>. Accessed 2009 September 7.
4. Smith B, Ashburner M, Ceusters W, Goldberg L, Mungall C, Shah N, Bard J, Eilbeck K, the OBI Working Group, Leontis N and others. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 2007;25:1251-1255.
5. Lynch CO. <http://lists.hl7.org/read/messages?id=162257>. 2009.
6. Scheuermann RH, Ceusters W, Smith B. Toward an Ontological Treatment of Disease and Diagnosis. Proceedings of the 2009 AMIA Summit on Translational Bioinformatics, San Francisco, California, March 15-17, 2009: American Medical Informatics Association; 2009. p 116-120.
7. 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(5):R46.
8. Schulz S, Johansson I. Continua in biological systems. *The Monist* 2007;90(4):499-522.