ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

River Flow Model of Diseases

Riichiro Mizoguchi[1], Kouji Kozakil[1], Hiroko Kou[1], Yuki Yamagata[1],
Takeshi Imai[2], Kayo Waki[2], Kazuhiko Ohe[2]

[1]ISIR, Osaka University, Ibaraki, Japan
[2]Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo, Japan

Abstract. This article discusses the ontological treatment of diseases in the framework of the Ontology for General Medical Science (OGMS). We aim to provide a definition of a disease that is more friendly to clinicians and propose a corresponding model of diseases. We define a disease as a dependent continuant constituted of one or more causal chains of clinical disorders. To clarify the ontological meaning of causal chains, we introduce two kinds of processes: a *cumulative continuous process* and a *non-cumulative process*. They are accounted for based on a new ontological theory of objects and processes. We then introduce the core ideas of a disease as causal chain and of clinical imbalance. We believe that the result can be considered as a concretization of the OGMS view of disease as disposition.

Keywords: disease ontology, causal chain, objects and processes, imbalance

1 Introduction

Recently, there has been a serious need for a consistent and ontologically sound medical vocabulary. The need is increasing as the tasks which need to be addressed by information technology in handling medical data become ever more demanding. In this situation, we believe that the Ontology for General Medical Science (OGMS) [1] based on BFO [3] and developed under the supervision of Barry Smith, is of considerable value. In OGMS, we find an excellent definition of a disease as: *a disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism*. This is a beautiful definition from a philosophical point of view. At the same time, however, it is not very friendly to clinicians because it lacks practicality. This reminds us of the reaction of engineers when they learn Smith's preferred account of function, which is also a beautiful one, as a type of disposition. However, engineers believe that function is something more real than a mere disposition. We believe that a domain ontology should be useful for both domain experts and ontologists.

We are not claiming that such a philosophically beautiful definition is useless to domain experts. Rather, we would like to try to develop another definition of disease that is more friendly to clinicians by concretizing the notion of disposition in keeping with the philosophy of vocabulary design used by OGMS.

This paper is organized as follows. We begin by analyzing the definition of disease in OGMS and explain our motivation in developing another definition of disease. Section 3 discusses our definition of disease. Section 4 provides an ontological theory of objects and processes to support this definition. Based on our definition, we propose in Section 5 a disease model that can be implemented on a computer.

2 Analysis of the Definition of a Disease in OGMS

Our concerns about the definition of disease in OGMS are as follows:

(1) Dispositions are introduced in the course of disease development in the human body. A disposition is a potentiality; on the OGMS view the realization of this potentiality takes the form of chains of physical/physiological changes in the human body. For OGMS currently, therefore, disease and disease course are distinguished; the latter is in a sense outside the former. We believe this use of 'disease' is counterintuitive to clinicians, and we thus propose a definition of disease that allows the disease to be placed within

the chain of events that is the disease course.

(2) To see what is missing from the current OGMS's approach, consider how a particular disease is identified in its terms. When explaining diabetes, for example, OGMS refers quite appropriately to an "elevated level of glucose in the blood". However, it provides an insufficient account of why the explanation of diabetes needs to mention "elevated level of glucose". What role does this elevated level play in diabetes itself? Why must "elevated level of glucose in the blood" be mentioned for diabetes but nothing else? It must be something specific to the disease of interest; that is, each realization of the disease must involve an entity of this sort. For OGMS, what plays the role is the disposition and the disorder (a certain disordered body part) in which this disposition inheres. We believe that the reference to elevated level of glucose points to the need for a further type of entity, which is included in our disease model.

We know there is a difficulty in defining such a type because it is not always definite for each disease, since it varies from one patient to another. Hence OGMS' use of disposition, which is a mere potentiality. In the case of latent diabetes, for example, there is no elevated level of glucose in the blood of the patient, though there is a disposition thereto. For latent diabetes, accordingly, we follow OGMS in recognizing the need for something other than just "elevated level of glucose in the blood". But we think that there is still something more that is required – something that is essential for each particular disease. In the case of diabetes, for example, this would be the *deficiency of insulin*, since this must have happened for all patients who suffer from diabetes. To tackle this issue, we draw on OGMS' notion of homeostasis and introduce the term 'disturbance of homeostasis' to explain what we see as the essential core of each disease. Disturbance of homeostasis can be caused through the concretization of a disposition, or it can be caused through some outside agency, for example through injury.

We agree with OGMS that a disease is a dependent continuant, and its definition is expected to address the following three conditions: (1) the existence of its pre-clinical manifestation, (2) the fact that it can cause another disease, and (3) variation in the disease course from patient to patient [1]. We try to find another definition of disease that satisfies these conditions.

3 What is a Disease?

Before going into discussion, we present some definitions of terms used in this paper. See [2] for details of event and process.

(1) A ***enacts*** B =def A is a continuant and B is an external process of A who/which participates in it as a whole in which it is maximal among participants who/which play the same role in the process. Examples: when you walk, you (not your legs) enact your walking, the motion of your legs is the internal process of your walking, something which you cannot enact.

(2) ***Event*** =def a non-dissective unitary entity in the temporal space. Examples include a conference, an arrival, etc. It must be dealt with as a whole in any case.

(3) ***Process*** =def a dissective non-unitary entity in the temporal space like walking, singing, etc. An event is constituted of processes, unless it is instantaneous.

(4) ***External process of A*** =def a process enacted by a continuant A.

(5) ***Internal process of A*** =def a process enacted by a part of A. Examples: In a walking process of A, leg motion is an internal process of A whose external process is the walking.

(6) ***Causal chain*** =def a chain of entities linked by causality. There can be a causal chain of disorder, causal chain of processes, causal chain of events, etc.[1]

3.1 Basic Strategy

We understand a typical disease as a dependent continuant satisfies the following: After it begins to exist, it enacts extending, branching, and disappearing processes before it disappears. Thanks to these processes, a disease can be identified as a continuant that is an enactor of those processes. Such an entity (a disease) can

[1] The topic of causality is here outside our scope.

change according to its phase while keeping its identity. OGMS defines such an entity precisely as a "disposition". Intuitively, however, it could be something related directly to a manifestation process of the disease rather than a disposition itself. At the same time, a disease should not be a process (occurrent) but a continuant. This is why defining a disease is difficult. Although the introduction of the notion of disposition is one way to solve this problem, for the reasons advanced above, disposition is a bit too far from what its manifestation process implies/suggests.

3.2 Definition

We can now define a disease as follows:

Definition 1: Disease

> A disease is a dependent continuant constituted of one or more causal chains of clinical disorders appearing in a human body and initiated by at least one disorder.

Then, what is a causal chain of disorders? Although it looks like a process, it is a dependent continuant. Some people might see that a causal chain of disorders is similar to a fall of water, river flow, fire of a forest, etc. We will show how a disease is a dependent continuant rather than a process in the next section. The following is an informal account of our view.

There are two kinds of processes:

(1) Cumulative continuous process[2]: a process that proceeds without completing the current process at every instant in time.

(2) Non-cumulative process: a process that proceeds by completing the current process at every instant in time.

Most processes, such as walking, eating, talking, etc., belong to type 2. What type 1 includes are falls of water, river flows, fires of a forest, etc. The key issue here is that those cumulative continuous processes are associated with continuants, called a waterfall, a river, a forest fire, etc. in these examples. This will be briefly discussed in section 4 based on the new theory of objects, processes, and events published in [2].

A causal chain is composed of one or more pairs of entities/events such as a causal event and an effect event, in which the latter has been caused by the former. The effect becomes another cause that causes another effect in the case of multiple-pair chains. What makes clinical causal chains special is that causal entities are usually still active when the effect entity has been caused. Therefore, the two entities overlap in temporal space. This shows that clinical causal chains belong to the type 1 process. In the case where the entities are continuants, by "an entity is active" we mean: it keeps its state as it is, so in the case of a disorder, the disordered organism still is the same disorder.

Let us examine how well a flowing river matches a causal chain of a disease. The river itself enacts branching, changing its shape, extension, diminishing, etc. In ancient times, when the river was initiated as a certain amount of water flowing, say, as a result the overflow from a lake or as a result of a heavy rainstorm, then the flow of the river is minimal. The overflow from a lake would correspond to an etiological disorder in a clinical causal chain. When the initial flow grows, the body of flowing water extends in length and is recognized as a river. After it has been born as a river (as a disease), then it extends further to another lake or to the sea. While extending, it branches (the branching perhaps causing the appearance of another disorder). Finally, it may dry up because of climate change (cure). Thus, the life of a river corresponds well to the life of a disease. Thus – in concordance with OGMS – both a river and a disease are continuants, though a river is an *independent* continuant but a disease (causal chain) is a *dependent* continuant which depends on a bearer, that is, a human being.

3.3 Discussion

On granularity: We do not specify any particular granularity of disorder and causal chains because we believe this should be flexibly determined according to the necessity of the description of each disease. Concerning the original cause, however, we have a policy that we should trace the causal chain back to the cell-level rather than to the genome-level. As far as we define diseases in general, granularity is not an issue, though it matters when we define a particular disease in the ontology.

We neither impose any specific time resolution on the causal processes so that we can

[2] The term *cumulative continuous process* was suggested by Barry Smith.

if needed include rapid processes such as fractures in our account. After receiving a strong external pressure, a bone undergoes a very quick destruction process resulting in fracture. The causal process can be captured by much finer time resolution than those involved ordinary pathological processes captured at the clinical level. Fracture can be dealt with by the disease model discussed in section 5.

On the distinction between a disease and a disorder: The distinction is shared with OGMS. Where OGMS defines disease (in brief) as a disposition, thus as a certain type of dependent continuant that is realized through pathological processes, we define a disease as a dependent continuant that enacts processes over pathological processes as causal chains of disorders towards a disorder(s). Disease course in OGMS is close to our definition; there, too, the disease course is a process.

4 Waterfalls and Rivers are Continuants

In order to support the above informal observation, we need to find a convincing ontological account of processes and objects. Due to space limitations, we here provide certain relevant passages from [2]. Then, we apply the discussion to support the definition of a disease as a dependent entity of a new type, different from a disposition and from a process.

Any change must be a change of something. This is already an argument against a 'pure process' view of reality, since we cannot conceive of processes without their material support. One might ask: what is a person over and above the sum of its internal processes? But what makes this sum worthy of consideration at all is that they constitute some kind of unity; the unity comes from the fact that there are other processes, its external processes which it enacts. Thus these questions make the mistake of focusing only on the internal processes of a person, whereas the external processes play an essential role in determining the identity of the object. Hence, rather than trying to characterize an object in terms of its internal processes (e.g., by identifying the object as the sum of those processes), we would rather say that an object is a unity which is what enacts its external processes. We could indeed say that the object is the interface between its internal and external processes: it is a point of stability in the world in virtue of which certain processes are characterized as internal and others as external. The issue of external vs. internal processes summarizes as "The water falls, but the waterfall doesn't fall". That is, what a waterfall is *doing* is not *falling the water* but migrating upstream as it carves its way into the rock.

Similarly, what a river enacts is not the water flow but change of the shape of its course. This is why we can consider a river as an object that has water flow as its internal process. Similarly, a causal chain as a flow of causality (propagation of causality) is an internal process of the causal chain which is a continuant that enacts branching, extension, and diminishing processes as its external processes. Although any disease has dynamic flows of the propagation of causality as its internal processes, it is the enactor of its cumulative continuous processes such as branching and extending its causal chain of disorders as its external processes.

5 A Model of Diseases

5.1 Core Causal Chain of a Disease

On the basis of the ontological definition of diseases, we build a computational model of diseases to make it easier to define particular diseases. In the following discussion, we divide diseases into two: (1) those whose etiological and pathological processes are well-understood and (2) other diseases, and we discuss them in turn.

Diseases of type 1 are identified by their inherent etiological/pathological process(es). Diseases of type 2 include so-called syndromes and are typically represented in terms of *criteria for diagnosis*. In this section, we deal with type 1 diseases first. Let us confirm that every disease of type 1 should have a clue for identifying it. That is to say, we should be able to find something like its so-called *main pathological/etiological condition(s)* that theoretically characterizes the disease to identify it. As stated above, this is what OGMS needs to include.

We know that diseases of type 2 necessarily employ *criteria for diagnosis* to identify them because of the lack of knowledge about their etiological/pathological processes. However, this does not mean it is excluded from our disease model as is discussed below, which we share

with OGMS.

We also need a formulation for organizing diseases in an *is-a* hierarchy in a disease model. According to our definition of a disease, this would consist of a causal chain(s) which consists of nodes and links, and hence a disease is represented as a Directed Graph. We can introduce an *is-a* relation between diseases using an inclusion relationship between causal chains as follows:

Definition 2: *Is-a* Relation between Diseases

Disease A is a superclass of disease B if all of the causal chains at the class level of disease A are included in those of disease B. The inclusion of nodes (disorders) is judged by taking an *is-a* relation between the nodes into account, as well as sameness of the nodes.

Definition 3: Core Causal Chain of a Disease

The causal chain of a disease included in the chains of all its subclass diseases is called the core causal chain of the disease. An example of the core causal chain in the case of (non-latent) diabetes is:

deficiency of insulin → elevated level of glucose in the blood.

Definition 3 helps us systematically to capture the necessary and sufficient conditions of a particular disease, which roughly corresponds to the so-called "main pathological/ etiological conditions". Fig. 1 shows the main types of diabetes constituted by corresponding types of causal chains. The most generic type in this example is *(non-latent) diabetes*, which is constituted by the chain:

deficiency of insulin → elevated level of glucose in the blood

The next lower subclasses include *type-I diabetes*, which is constituted by:

destruction of pancreatic beta cells → lack of insulin I in the blood

→ deficiency of insulin → elevated level of glucose in the blood

and *steroid diabetes*, which is constituted by:

long term steroid treatment → … → deficiency of insulin

→ elevated level of glucose in the blood

If a doctor wanted to have a hierarchy representing diabetes-caused blindness, then it would be:

deficiency of insulin → elevated level of glucose in the blood → … → loss of sight

Due to space limitations, we omit here the discussion about the cases of resistant peripheral receptors which are also covered by our model.

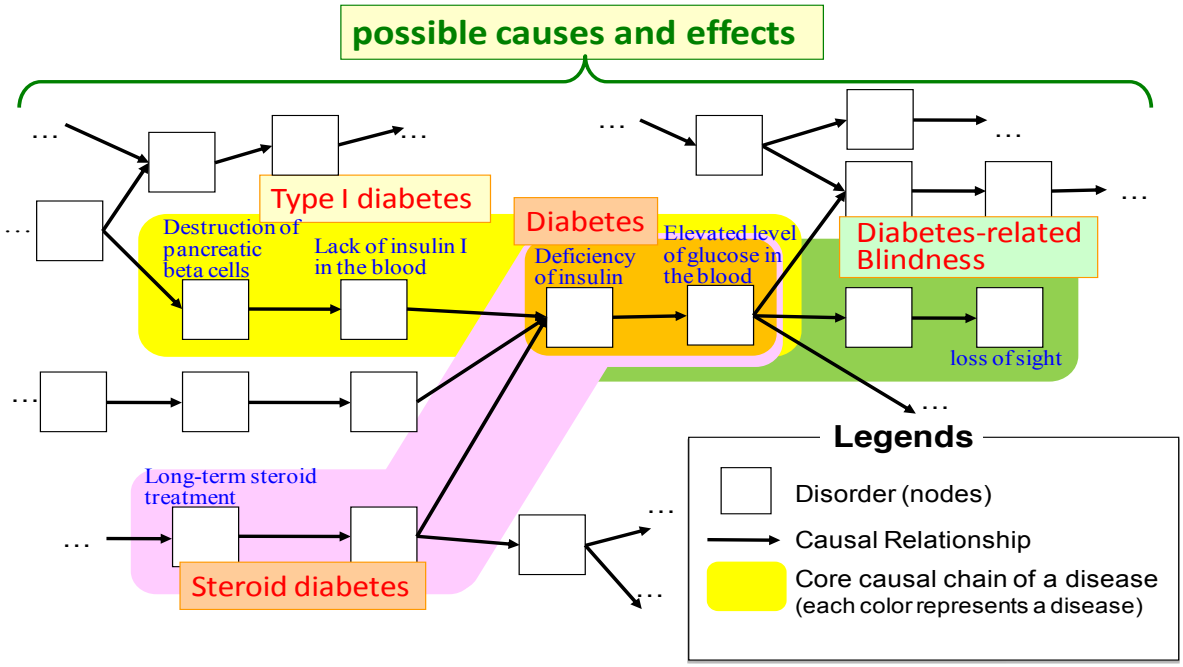

Figure 1. Types of diabetes constituted of causal chains.

Although we explain the disease model using Type 1 diseases as example, it is applicable also to Type 2 diseases thanks to the flexibility of granularity and degree of being "well-understood". These two kinds of flexibility can be exploited according to each disease under consideration. In the case of diseases of Type 2, we could employ an "unknown" causal node linking to just a few of those symptoms that are typically observed in the case of the syndrome under consideration. Note that this model can capture a seemingly isolated symptom by combining it with an unknown cause to form a causal network. It also captures diseases with multiple causal chains.

One might suspect that this model cannot cover a phenomenon such as obesity due to the too large variety of associated causal chains, so that the classification according to Definition 2 above does not make sense. However, our model does cover obesity successfully, since it accepts multiple causal chains. Because those causal chains are not essential to obesity, unlike diabetes, they are not included in the core causal chain. Hence we do not have to classify obesity according to those causal chains. Instead, our ontology tool, HOZO [5], used for implementing the medical ontology, has a function to dynamically generate *is-a* hierarchies of diseases according to a perspective given by users [4]. Although it has some limitations, this function allows us to leave diseases in a rather flat structure if appropriate, and users classify them afterwards using the function.

Our model also can distinguish, for example, between diabetes with blindness and diabetes-driven blindness by specifying the core causal chain that is focused upon. In summary, the disease model yielded by the definition of disease proposed in section 3.2 above (Definition 1) covers quite a wide range of diseases. In fact, we have built models of 6051 diseases from 12 different divisions in our ontology, which shows the expressive power of our disease model.

5.2 Imbalance

Now, we specify the disease model discussed above by restricting diseases to deal with those of Type 1. We can introduce a mechanism to effectively model diseases of this type.

In OGMS, it seems to us that a disposition to diabetes inheres in <deficiency of insulin>. As we can easily see, <deficiency of insulin> is a physical state which, in principle, can be detected. Then, the issue is how to capture such an entity in a computational model? We have come up with the idea of *disturbance of homeostasis*. By homeostasis, we mean the same as is described in OGMS [1, p. 117]. For each parameter participating in homeostasis, there must be the notion of balance and regulation functions.

We can understand the notion of balance by introducing a performable operation (**supply**) and a required operation (**demand**). In the case of diabetes, the former is the performed amount of the insulin operation and the latter is the required amount of the insulin operation. In a normal case, the difference between the two amounts is within a certain range, that is to say, "balanced". In an abnormal case, on the other hand, an imbalance (deficiency of insulin) occurs, which can be a disposition to the initiation of the pathological process of diabetes. On the basis of the above discussion, we define a concretized disposition as follows:

Definition 4: Clinical Imbalance

> Clinical imbalance is a local phenomenon of homeostasis in the human body and is defined as a state where the difference between the amounts of supply and demand is out of the range specified for the parameter under consideration.

Precisely speaking, this balance model is represented by four nodes (see Fig. 2): balance between supply and demand, performed amount of the operation (supply), possible maximal amount of the operation, and required amount of the operation (demand). Supply can change adaptively in response to changes in the amount of demand, but only up to the maximal amount. If demand exceeds the maximal amount, a clinical imbalance occurs.

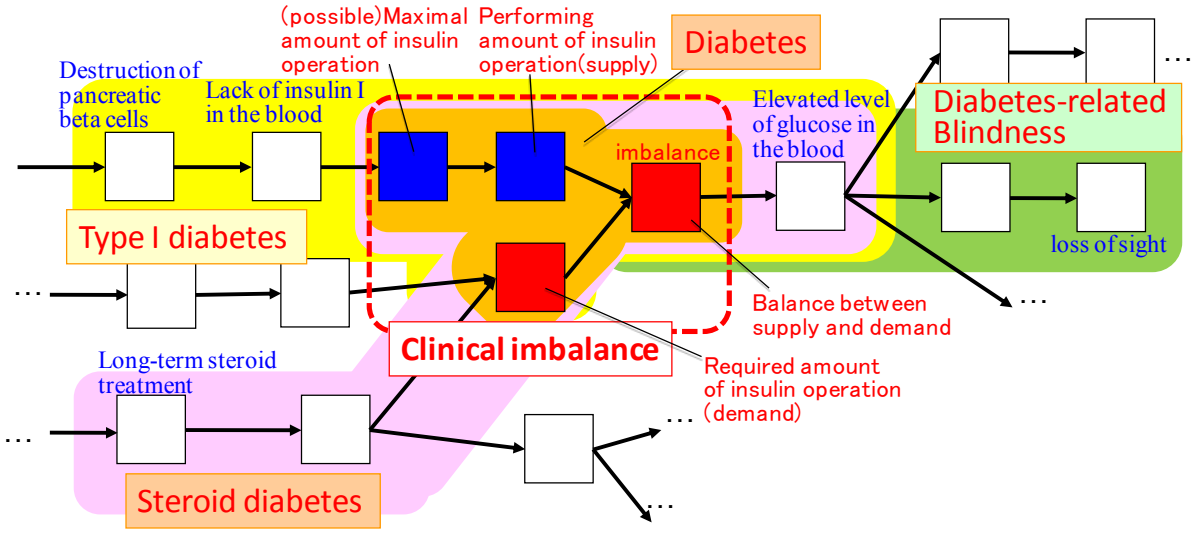

Figure 2. A representation of the clinical imbalance model

This discrimination of possible causes is critical to the proper understanding of diseases. To exploit the notion of clinical imbalance, we need some quantized generic values: **small**, **medium**, **large**, and **very large**[3]. By **medium**, we mean the quantity that a patient needs in everyday situations. By **small**, we mean the quantity that a patient needs in a very calm or inactive situation. By **large**, we mean the quantity that a patient needs in a stressful or active situation but that can be coped with by a normal regulation function. By **very large**, we mean the quantity that cannot be coped with by a normal regulation function. The above four nodes, except *balance*, take these four values. Due to space limitation, we have to omit an explanation of how to use the notion of clinical imbalance. Instead, we discuss the characteristics of our disease model with clinical imbalance as a factor, employing the mentioned four qualitative values.

(1) The model can correctly capture diseases such as latent diabetes where, in OGMS terms, the relevant disposition is *not realized* at the level of clinical manifestations. In the case of latent diabetes, although maximal amount of insulin supply is **medium**, that is, smaller than those of healthy persons, the supply of insulin operation can cope only with a demand less than **large**. Therefore, no imbalance occurs while the patient is going about her normal daily activities. However, the fact that the maximal amount is smaller than **large** shows that the patient is said to suffer from latent diabetes. If the demand for insulin is increased for some reason and becomes greater than the maximal amount of insulin supply, then imbalance occurs, and the diabetes is no longer latent.

(2) We evaluated the expressive power of the model by representing diabetes, ischemic heart diseases, infectious diseases, and osteoporosis and found it worked satisfactorily. For example, fracture caused by osteoporosis is modeled using **medium** value for demand to resist the normal pressure given as an external cause and **small** value for supply to resist external pressure. While bones of normal people can stand such external pressure, patients that suffer from osteoporosis cannot, which the imbalance model clearly explains.

(3) These four qualitative values work well to represent each particular disease whose instances share the same threshold values to quantize real values, though it does not make sense to compare them across different diseases. Because our goal is defining each particular disease rather than diagnosis, we do not need concrete thresholds or ranges of their values.

[3] Although it is explained in terms of the *demand* side in the following, it also applies to the *supply* side in a similar way.

(4) The parameter(s) chosen in the model should be dependent on the particular disease under consideration but causes no problem, since it is what the medical experts think essential for capturing the disease.

6 Concluding Remarks

We have discussed a definition of a disease friendly to domain experts based on a new ontological theory of objects and processes. We conducted a small informal evaluation by asking seven medical doctors with different expertise who are totally unfamiliar to ontology and do not know the authors which definition they like better among the two definitions and learned all selected our definition, which suggests our definition is friendly to them more than that in OGMS.

The definition enables us to understand a disease as a dependent continuant constituted of a clinical causal chain(s). We also discussed a model of a disease that allows the definition to be implemented. In the model, we defined a core causal chain of a disease with the idea of clinical imbalance which can be considered as the concretization of disposition to a disease. With this approach, we believe that we have improved OGMS. We have been developing a medical ontology for the last four years on the basis of our definition and model of diseases [4]. As of April 12, 2011, a total of 6051 diseases have been defined in the medical ontology by 12 clinicians, and these definitions are currently being refined. Currently, we have no concrete connection with activities conducted outside Japan, but we are open for collaboration with ventures as DO [6], OBO[7] as well as with OGMS.

Acknowledgement

A part of this research is supported by the Japan Society for the Promotion of Science (JSPS) through its Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program). The authors are grateful to all the reviewers, who gave us valuable and constructive comments.

References

1. Richard H. Scheuermann, Werner Ceusters, and Barry Smith, Toward an Ontological Treatment of Disease and Diagnosis, Proceedings of the 2009 AMIA Summit on Translational Bioinformatics, San Francisco, CA, 2009. pp. 116-120 (2009)
2. Antony Galton and Riichiro Mizoguchi, The water falls but the waterfall does not fall: New perspectives on objects, processes and events, Applied Ontology 4(2), pp. 71-107 (2009)
3. BFO: http://www.ifomis.org/bfo/
4. Riichiro Mizoguchi, Hiroko Kou, Jun Zhou, Kouji Kozaki, Ken Imai and Kazuhiko Ohe, An Advanced Clinical Ontology, Proc. of International Conference on Biomedical Ontology (ICBO), Buffalo, NY, June 24-26, 2009, pp. 119-122 (2009)
5. HOZO: http://www.hozo.jp/
6. DO,http://do-wiki.nubic.northwestern.edu/index.php /Main_Page/
7. OBO: http://www.obofoundry.org