

# The Problems of Realism-Based Ontology Design: a Case Study in Creating Definitions for an Application Ontology for Diabetes Camps

James C. Schuler, BSc<sup>1</sup>, PhD student and Werner M. Ceusters, MD<sup>1</sup>

<sup>1</sup>Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA

## Abstract

*A requirement of realism-based ontology design is that classes denote exclusively entities that exist objectively in reality and that their definitions adhere to strict criteria to ensure that the classes are re-usable in other ontologies while preserving their ontological commitment. Building realism-based ontologies is therefore quite challenging and time-consuming, demanding considerable training. Although the top-level in the form of the Basic Formal Ontology (BFO) is worked out very well, and also the upper levels of certain domains, there is still a disconnect with the bottom-up or middle-out approach which is typical, and more practical, for application ontologies. Using the development of an application ontology for diabetes management in diabetes camps as an example, we present an overview of problems trainees in realism-based ontology design can be confronted with and offer some guidelines on how to deal with them in case no ideal solution is available.*

## Introduction

Realism-based ontology design<sup>1</sup> is often criticized as being too difficult<sup>2</sup>. That it is difficult indeed is witnessed by the many ontologies that attempt adherence to the principles of Ontological Realism<sup>3</sup> but then fail at one or other stage and do not become accepted as an Open Biomedical Ontologies (OBO) Foundry ontology<sup>4</sup>. There are at least two reasons for this difficulty. One is that ontological commitment in realism-based ontologies is exclusively towards what has (or once had) objective ‘mind-independent’ existence<sup>3</sup>. This includes entities and their types (the latter called ‘universals’), for example organisms and cells, and certain combinations thereof (called ‘configurations’; entities, universals and configurations all being ‘portions of reality’) such as all living and deceased diabetes patients. The second one is that definitions for representational units in realism-based ontologies must adhere to very strict criteria: each one must be defined by means of an Aristotelean definition which states all necessary conditions that are also jointly sufficient and satisfiable for some entity to be an instance of the class denoted by the representational unit<sup>1</sup>. This is extremely important for *reference ontologies*, i.e. ontologies that are intended to be comprehensive for a given domain and that consist of representational units that can be re-used in other ontologies while preserving their ontological commitment to the original entities in reality. Representational units in the form of terms – in contrast to, for example icons or numerical identifiers – should also have general face validity independent of context<sup>5</sup>.

Concept-based ontologies such as SNOMED CT<sup>6</sup>, in contrast, require mere adherence to formal logical criteria and an ontological commitment not to reality, but to a ‘universe of discourse’. The advantage of concept-based ontology design is that the notion of ‘*concepts as shared meanings*’ is perceived as more intuitive and closer to the language used in the domains that are modeled than the quite ‘abstract’ notions of ‘universals’ and ‘portions of reality’ as employed by realists. While, for example, there is no place for terms such as ‘prevented abortion’ and ‘absent nipple’ in realism-based ontologies, they can be allowed in concept-based ontologies as, from such a perspective, it does not involve rocket-science to grasp their meaning. Therefore, metaphorically, conceptualists can ‘*see*’ both *absent nipples* and *persons with absent nipples*, while realists can in that context only see *persons with absent nipples*. While the challenge for conceptualists is not to misclassify prevented abortions and absent nipples as special kinds of abortions, resp. nipples, the challenge for realists is to find ways for formally describing in what way persons with absent nipples are different from persons with prevented abortions without resorting in these descriptions to absent nipples and prevented abortions<sup>7</sup>. Such considerations are very hard for trainees without a solid education in philosophy.

Whereas strict adherence to the realism-based principles is an absolute requirement for reference ontologies<sup>3</sup>, one might wonder whether that must also be the case for *application ontologies*, i.e. ontologies which are designed to accomplish some specified local task or application within a specific context. Application ontology development usually follows a bottom-up or middle-out approach: fundamental terms in the application domain are addressed first before moving up to more generic and down to more specific terms. One could argue that for such ontologies it is sufficient to introduce terms that have face validity only within the interface terminology<sup>8</sup> of the intended application. One might even be tempted to simplify definitions to what is relevant for the logical semantics of the underlying

application as advocated by concept-based ontologies. With less stringent criteria advanced ontological and philosophical analyses can be avoided, thus leading to quicker development times. The drawback is, however, that such ontologies are more prone to contain errors<sup>9, 10</sup>, and that formally underspecified terms with local face validity hamper their re-usability in other semantic applications so that data collected through them cannot automatically be merged and integrated with data coming from other applications, even if they cover the same domain. Therefore, it is still worthwhile to define the mid-level application domain representational units of application ontologies in terms of upper-level units from realism-based ontologies. Unfortunately, here lies an additional level of complexity for trainees in realism-based *biomedical* ontology design: whereas the top-level in the form of the Basic Formal Ontology (BFO) is worked out very well<sup>1</sup>, there is a shortage of adequate mid-level representational resources that follow realism-based principles strictly, perhaps with the exception of the Ontology for General Medical Science (OGMS)<sup>11</sup>.

In this paper, we demonstrate the sort of problems this introduces for trainees on the basis of a use case: an application ontology for glycemic control of attendees of diabetes camps. We offer some guidelines on how to deal with them in case no ideal solution is available. The goal is to identify and prioritize issues that needs to be addressed by experts to make realism-based ontology development more accessible.

### Background

The American Diabetes Association (ADA) holds many residential camps (‘diabetes camps’) for children age 7 to 18 with diabetes throughout the United States. These are typical summer camps in the sense the campers do all the typical camp activities over the course of a week, including activities such as arts and crafts, singing songs, building campfires, performing skits, rock climbing, archery, athletics, eating meals in a dining hall, and participating in an evening program revolving around a theme for the week. The atypical aspect of these camps is that all of the campers, and much of the staff and counselors, have diabetes. These camps have been held for over 90 years<sup>12</sup>.

Managing diabetes is a challenging task for many people with diabetes in a normal day-to-day scenario<sup>13</sup>. Adding the spectrum of high activity and a different routine than normal at camp only increases the difficulty of diabetes management. To keep campers safe, the management of the campers’ diabetes is achieved by means of teams composed out of medical professionals (e.g., physicians, nurses, and pharmacists, and medical professionals in training such as medical, nursing, pharmacy, and physician assistant students), the counseling staff, and most importantly, the campers themselves. Presently at camp, medical professionals working with the campers record key data on paper forms which describe the camper’s glycemic control and management (Figure 1).

CAMPER GOALS:				AFTERNOON		DINNER	EVENING	BED	NIGHT		
NOTES:				Time	BG	Time	BG	Time	BG	Time	BG
				Notes	Actions	Carb Count	Notes	Actions	Carb Count	Notes	Actions
				Time	BG	Carbs	Food Ins	Time	BG	Time	BG
				Notes / Actions	Notes / Actions	Notes / Actions	Notes / Actions	Notes / Actions	Notes / Actions	Notes / Actions	Notes / Actions
MONDAY	BREAKFAST	MORNING	LUNCH								
TUESDAY											
WEDNESDAY											
THURSDAY											
FRIDAY											

  

<b>CAMPER MED PROFILE:</b>	<b>Low Symptoms:</b>	<b>Activity Level (circle):</b>	
Allergies _____	_____	Mild Moderate Active	(cabin)
Celiac Dx? Y <input type="checkbox"/> N <input type="checkbox"/>	Hx of Seizures? Y <input type="checkbox"/> N <input type="checkbox"/>	Recent Illness? Y <input type="checkbox"/> N <input type="checkbox"/>	(age)
	Last seizure date: _____	Year at camp: _____	(last, first)

**Figure 1:** The current paper-based medical records used at Camp Aspire

The tracking of this data allows the medical professionals to perform analysis and make adjustments to the various aspects of care which determine glycemic control<sup>12</sup>. Moreover, the tracking on records is important for communication between staff members, and engages staff to provide continuous safe and high-quality diabetes management.

The current state of medical record keeping and tracking of vital healthcare data at diabetes camp is archaic and thus ripe for innovation. The data tracked on paper medical records include, but are not limited to: time of observation, blood glucose level, carbohydrates consumed in grams, doses of insulin, basal insulin delivery rate, and urine/blood ketone level. The paper records are cheap, easily modifiable, and relatively secure. However, there are many downsides to their use. For example, paper records only modestly facilitate tracking and analysis of data and are ill-suited to detect patterns indicating potentially dangerous developments<sup>14</sup>. Furthermore, the records can often be unreadable, and unavailable in a reasonable time frame in an emergent situation. Finally, there is no standard of use of paper records, or fields present in those records throughout the camps run by the ADA. Indeed, no two camps use the same forms for tracking of data (Figure 2). While each camp has its own unique forms, i.e., there is a standard within a camp, there is no standard between camps, as, for example, certain entities are referred to by different terms and possible ‘allowed’ values as instances. This lack of standard hinders research about diabetes management at all camps as data are not easily collated and analyzed what hampers the advancement in elucidation and understanding of the factors relevant to diabetes management in a camp setting and potentially places campers at risk for severe emergency events related to their diabetes. It is unknown to what extent an electronic medical record system implemented at a diabetes camp would impact the ease of recording and tracking data, elucidation of patterns in the data, and thus would increase the safety, care, and efficiency and efficacy of diabetes management at camp. With an ontology-based mobile app for diabetes care in such camps, this question might be answerable.

*Allergies/Warnings:* \_\_\_\_\_

**CAMP INSULIN LOG**

Name \_\_\_\_\_ Cabin/Session \_\_\_\_\_ / \_\_\_\_\_ Date of Diagnosis \_\_\_\_\_  
 Age \_\_\_\_\_ Weight \_\_\_\_\_ Target # (D) \_\_\_\_\_ (N) \_\_\_\_\_  
 Home Dose: (B) \_\_\_\_\_ (L) \_\_\_\_\_ (D) \_\_\_\_\_ (HS) \_\_\_\_\_

**Fast Acting: Novo - H - R Inter/Long Acting: NPH - Levemir - Lantus**

**(Time given):** \_\_\_\_\_ **Meal Plan #** \_\_\_\_\_

DATE	BLOOD SUGAR/KETONES						INSULIN ORDER			INSULIN GIVEN			
	12	2	B	L	D	HS	Time	Dose	MD	Time	Dose	Site	Staff Initials
							B						
							L						
							D						
							HS						
							B						
							L						
							D						
							HS						
							B						
							L						
							D						
							HS						
							B						
							L						
							D						
							HS						

**Figure 2.** The paper-based medical record template used at a different ADA camp than Camp Aspire.

## Methods

One option for the final assessment of students in the 500-level 3-credits course ‘Introduction to Biomedical Ontologies’ co-organized by the Departments of Biomedical Informatics and Philosophy of the University at Buffalo is the development of an ontology. The first author of this paper, a student in this course, being a regular collaborator at diabetes camps, envisioned the benefits of a small custom-made ontology-based electronic medical record app for glycemic control in such camps. It should be able (1) to support collecting, recording, presenting and managing data about the campers’ so that (2) the medical professionals would be able to analyze and interpret the data collected more easily. To keep it manageable, it would (3) allow collection of the data fields *pertinent* to the diabetes management of adolescents in a camp setting as, thus covering what is currently recorded on paper. This custom, micro electronic medical record would have less functionality than typical electronic medical records used in traditional inpatient or outpatient settings, but was hypothesized to have high value for use at diabetes camps. For reasons of reusability and comparability of data and integration with other relevant technologies, another design criterion envisioned would be (4) to keep the ontology – henceforth called the ‘Diabetes Camp Ontology’ (DCO) – underlying the system compatible

with the growing number of realism-based ontologies<sup>1</sup>. The application would, for instance, be able to deal with the following scenario from camp which forms the core of the relationship between a camper and the medical staff, and the majority of what is currently recorded in the paper records: ‘*Rob Smith checks his blood sugar when he wakes up on Monday. His sugar is 208. Rob’s medical staff for the week, Jim, records his blood sugar and based on this blood sugar, calculates Rob needs to take 2.8 units of insulin, and Rob takes his insulin. Immediately afterward, Rob Smith goes to the dining hall to eat breakfast. At breakfast he eats 65 grams of carbohydrates. Based on the number of carbohydrates Rob eats, Jim calculates Rob should deliver 4.3 units of insulin, so Rob does.*’ Such a scenario would be recorded in the paper records as displayed in Figure 3, a part of the form displayed in Figure 1.

In a first step, the student collected from the available paper record templates of various diabetes camps all the data elements used for the campers’ follow ups. In a second step, he inspected existing ontologies for the presence of representational units denoting types of entities in terms of which the entities represented by the terms that were collected could be described following the provisions of the Relation Ontology<sup>15</sup>. Whenever in a feeder ontology a representational unit was found that denotes an entity also referenced by one of the collected data elements, that representational unit became part of the DCO, otherwise, a new representational unit was defined.

NOTES:

BREAKFAST		M		
MONDAY	7:30 am	208	Time	Bi
	65		Time	Bi
	2.8 ins	4.3 ins	Time	Bi
	Notes / Actions		Time	Bi
Temp Basal				

Figure 3. Part of a diabetes camp form

The resulting ontology and the documentation thereof – a short paper and presentation were part of the assignment – was then inspected by the instructor for adherence to realism-based quality principles<sup>1,3</sup> and the total effort graded. At that stage, it became clear that failure to adhere was not only because of gaps in understanding on the side of the student, but also because of gaps in the available resources. The definitions and terms of the application ontology were therefore afterwards in a short turn-around modified and improved in discussion with the instructor to the extent possible and then analyzed together with the student’s arguments for his decisions with the goal (1) to find out what problems this student – and perhaps, to be determined through future studies, students in general – encountered with the application of realist principles, (2) to determine the root causes for these problems, and (3) to propose ‘easy fixes’, if any at all, that can be applied to definitions so that the corresponding classes, despite coming from an application ontology, can be more reliably, though not necessarily automatically, compared with relevant ones in other application ontologies within the same or a closely related domain

## Results

Building this application ontology turned out to be a difficult task and several issues were encountered. The biggest problem was not only to find adequate ontologies, but also adequate terms within them. The Basic Formal Ontology (BFO)<sup>1</sup> as a top-level ontology, and the Information Artifact Ontology (IAO)<sup>16</sup>, the Ontology for General Medical Science (OGMS)<sup>11</sup>, the Ontology for Biomedical Investigations (OBI)<sup>17</sup>, and the Ontology of Medically Related Social Entities (OMRSE)<sup>18</sup> as high-level domain ontologies turned out to be good choices despite the wide range of incompatible options for existing representational units (RU) to choose from. Whereas the ‘middle-out’ development approach, i.e. starting with terms that need to be part of the interface terminology of the application in line with the envisioned scenarios, provided a good basis, way more terms than originally anticipated needed to be included as well and this for purposes of (1) crafting adequate definitions – a challenge in its own right, (2) providing additional functionality to the app as compared to what is possible with paper forms and (3) the ones required to drive the logic of this app. A major challenge was to identify the exact ontological types of the entities that figured in the scenarios we had in mind, a problem which not only seems to be a problem for students and trainees in realism-based ontology development, but also for the experts. For the purposes of this paper, detailed analyses are limited to representational units and definitions related to camps and the roles played by participants (Table 1) and to blood glucose levels and insulin administration (Table 2). For each RU, a textual definition precedes a semi-formal definition. The latter uses the formatting principles of the Relation Ontology<sup>15</sup>: (1) each representational unit starts with the acronym of the ontology in which it is included, (2) SMALL CAPS is used for types, (3) *italic bold font* for formal relationships. This formatting is, where important for correct understanding what we are referring to, also used in the discussion section.

## Discussion

Finding relevant domain ontologies. Several projects report to have attempted the creation of an ontology for diabetes, its treatment, its complications or its diagnosis often with the eventual end goal to create clinical decision support systems to aid in numerous diabetes related issues for both patients with diabetes and healthcare providers<sup>19-25</sup>.

**Table 1.** Terms and definitions related to camps and roles played by participant

<b>Class and definitions</b>
<p>DCO:CAMP</p> <p>D1a: an organization which provides people (often children) specialized facilities and activities, often related to a central theme or goal, and often providing overnight accommodations.</p> <p>D1b: OBI:PLANNED PROCESS which (1) <i>realizes</i> the BFO:ROLE of some OBI:ORGANIZATION at some BFO:SITE which comprises overnight accommodations, and (2) <i>has process parts</i> (2a) in which <i>instances of</i> OBI:HOMO SAPIENS <i>participate</i> and (2b) some of which are of a sort as described in <i>instances of</i> IAO:OBJECTIVE SPECIFICATION describing some central theme or goal.</p>
<p>DCO:MEDICALLY-FOCUSED CAMP</p> <p>D2a: a camp where an instance of some disease inheres in all instances of attendees, and whose mission often includes creating a sense of normalcy for the attendees and teaching self-management of the disease, while providing necessary and critical medical care.</p> <p>D2b: DCO:CAMP which (1) <i>has attendees of</i> OBI:HOMO SAPIENS <i>HS</i> all of which <i>are inherers of</i> some <i>instance of</i> some OGMS:DISEASE <i>D</i>, (2) <i>has process parts</i> some of which are of a sort as described in <i>instances of</i> IAO:OBJECTIVE SPECIFICATION describing the objectives to create a sense of normalcy for <i>HS</i> and to teach self-management of <i>D</i>, and (3) <i>has process parts</i> some of which provide necessary medical care to <i>HS</i>.</p>
<p>DCO:DIABETES CAMP</p> <p>D3a: medically-focused camp, where the specific which inheres in all instances of attendees is diabetes.</p> <p>D3b: DCO:MEDICALLY-FOCUSED CAMP which <i>has attendees of</i> OBI:HOMO SAPIENS all of which <i>are inherers of</i> some <i>instance of</i> diabetes.</p>
<p>DCO:CAMP ASPIRE</p> <p>D4b: DCO:DIABETES CAMP organized by the OBI:ORGANIZATION called ‘<i>American Diabetes Association</i>’.</p>
<p>DCO:DIABETES CAMP CAMPER ROLE</p> <p>D5a: a patient role borne by a human being realized by participating in the events held by some diabetes camp.</p> <p>D5b: OMRSE:PATIENT ROLE which <i>inheres in</i> OBI:HOMO SAPIENS which <i>are attendees of</i> some DCO:DIABETES CAMP.</p>
<p>DCO:DIABETES CAMP MEDICAL STAFF ROLE</p> <p>D6a: a health care provider role borne by some human being and realized by managing or directing of management the diabetes care of a diabetes camp camper.</p> <p>D6b: OMRSE:HEALTH CARE PROVIDER ROLE which <i>inheres in</i> OBI:HOMO SAPIENS and which <i>is realized by</i> OBI:PLANNED PROCESS(es) of managing or directing the diabetes care of DCO:CAMPER(s).</p>
<p>DCO:CAMPER</p> <p>D7a: human being which is bearer_of some diabetes camp camper role.</p> <p>D7b: OBI:Homo Sapiens which <i>is bearer_of</i> DIABETES CAMP CAMPER ROLE.</p>

The existence of them may raise the question whether yet another diabetes ontology is needed? Closer inspection reveals that only few have potentially pertinent terms and relations<sup>26,27</sup>. Unfortunately, rarely do these projects follow the realism-based guidelines and only one<sup>28</sup> expresses interest in following guidelines to become a part of the OBO Foundry. While all are meaningful work within a concept-based approach and for their own specific uses, none encapsulate the reality of diabetes care and management in a camp setting. Therefore, for our purposes, a new realism-based ontology with high potential for integration with OBO Foundry compliant ontologies to address the specific needs of serving as a model of reality of caring for children with diabetes in a camp setting is indeed needed.

What terms to include? In line with the purposes the ontology has to serve, a first type of representational units to be included are those that represent entities to which the terms from the interface terminology of the application stand in an *aboutness* relation<sup>16</sup>. This includes entities referenced by the terms on the paper forms used in diabetes camps, as well as the ones for which the application might offer some possible values. How clear cut this might seem, it isn’t in light of the specific ontological commitment of the BFO and the ontologies that intend to be compatible with it. An example is what is denoted by the abbreviation ‘BG’ on the form in Figure 1, i.e. blood glucose levels, and what is expected to be entered as values to denote a specific blood glucose level. One could allow just numbers such as ‘98’ and ‘289’, standing for ‘blood glucose level of 98 mg/dl (milligrams per deciliter)’ and ‘blood glucose level of 289 mg/dl’ resp. But does BFO commit to the existence of numbers with objective, ‘mind-independent’ existence? This is to say, if we write ‘98’, does this term denote some *x* in reality? If so, would that *x* be *an instance of* the type number?

**Table 2.** Terms and definitions related to blood glucose and insulin administration

<b>Class and definitions</b>
<p>DCO:BLOOD GLUCOSE LEVEL</p> <p>D8a: A physical quality which inheres in blood by virtue of the concentration of glucose in the serum.</p> <p>D8b: BFO:QUALITY which <i>inheres in</i> a portion of blood by virtue of the concentration of glucose in the serum.</p>
<p>DCO:BLOOD GLUCOSE READING</p> <p>D9a: a scalar measurement datum that is the specified output of measuring glucose concentration in serum, often displayed on a glucometer.</p> <p>D9b: IAO:SCALAR MEASUREMENT DATUM that <i>is the specified output</i> of a OBI:MEASURING GLUCOSE CONCENTRATION IN BLOOD SERUM, often displayed on an OBI:GLUCOMETER, and which <i>is about</i> some DCO:BLOOD GLUCOSE LEVEL</p>
<p>DCO:INSULIN PUMP</p> <p>D10a: a device which stores insulin and participates in a process of subcutaneous injection with or without human intervention, according to some pump setting datum.</p> <p>D10b: OBI:DEVICE in which <i>inheres</i> the BFO:FUNCTION(s) (1) to store insulin and (2) to participate in instances of OBI:SUBCUTANEOUS INJECTION, according to some DCO:PUMP SETTING DATUM</p>
<p>DCO:INSULIN PEN</p> <p>D11a: a device which stores insulin and participates in a process subcutaneous injection with human intervention.</p> <p>D11b: OBI:DEVICE in which <i>inheres</i> the BFO:FUNCTION(s) (1) to store insulin and (2) to participate in instances of OBI:SUBCUTANEOUS INJECTION carried out by an instance of OBI:HOMO SAPIENS.</p>
<p>DCO:INSULIN ADMINISTRATION SITE</p> <p>D12a: a part of the host body and entrance point where insulin is administered</p> <p>D12b: BFO:SITE which <i>is part of</i> a DCO:CAMPER through which insulin is administered</p>
<p>DCO:INSULIN INJECTION SITE</p> <p>D13a: an insulin administration site that is used for the injection of insulin</p> <p>D13b: DCO:INSULIN ADMINISTRATION SITE that is used for injection of insulin</p>

Or would that  $x$  be a type itself so that there are entities which *are instances of* that  $x$ . Are then 98 apples, a heart rate of 98 beats per minute and a blood glucose level of 98 mg/dl all instances of  $x$ ? Trainees might have these questions, yet find no answers in available documentation, or perhaps even more confusing, different positions which nevertheless all fall under a realist perspective. It is of course here that pragmatism comes into play by leaving out anything not directly relevant to the application while still staying under the realist perspective. So it is documented that BFO commits to entities on the side of the patient that are of types such as ‘blood glucose level of 98 mg/dl’ – they existed before anybody knew about glucose in blood, how to measure it, etc. – and all these entities are also of the type ‘blood glucose level’. Because blood glucose level is a quality, it allows convenient reasoning and accurate representations of reality at instance levels. For example, consider an instance of camper, Rob Smith. Rob Smith’s blood glucose level exists, and is an instance of blood glucose level for all the time during which Rob Smith exists. At some time,  $t_1 = 7:30$  am, Rob Smith’s blood glucose level instantiates a blood glucose level of 208 mg/dl. At some later time,  $t_2 = 10:30$  am, Rob Smith’s blood glucose level instantiates a blood glucose level of 115 mg/dl. Software can deal appropriately with these issues without the need to overload the ontology.

Should one include terms such as ‘hyperglycemia’ or ‘hypoglycemia’? These are medical terms describing blood glucose which often have strict definitions and cut-off values in certain settings and situations which often dictate action or treatment protocols. For example, in the emergency department of a hospital if a patient’s blood glucose is described as ‘hypoglycemia’ (physiologically low), it might prompt a provider to obtain some sugary food or drink for the patient to raise the blood glucose level back to ‘euglycemia’ (physiologically normal). Kids attending diabetes camp may have different blood glucose levels which they describe as ‘high’, ‘normal’, or ‘low’, a statement often referring to a blood glucose level in some range of blood glucose levels which is decided by the child’s physician. Furthermore, ‘High glycemia’ and ‘hyperglycemia’ do not always coincide, and neither ‘normal glycemia/euglycemia’, and ‘low glycemia/hypoglycemia’. One might be tempted to include terms such as ‘within target range blood glucose’, ‘above target range blood glucose’, and ‘below target range blood glucose’. But then, targets might be different for each patient. Thus in two distinct kids might inhere blood glucose levels which are of exactly the same determinate type, yet one could be within target, while the other one wouldn’t be. The issue here is that such terms when used on the form of some patient are not ‘just’ *about* the blood glucose level of that individual patient, but also

*about* the population of human beings or about the disease course of that patient as defined in the OGMS<sup>11</sup>. When it is decided to use these terms, it is mandatory that the ontology contains a second type of representational units, i.e. those which are required to provide adequate ontological definitions<sup>29</sup>. But there is not always the need for these terms to become also part of the interface terminology of the application.

The third type of representational units to be included are those required for the app to add additional functionality as compared to what is possible with paper forms. Thus it is critical for the functioning at diabetes camp and the representation thereof in the ontology for the app that certain roles be defined and used in precise manner. For example, when reviewing the records of diabetes management at camp, it is important to know who ordered what dose of which medication, as certain roles confer certain powers to individual persons to make certain decisions regarding the diabetes management of campers. Consider the following: a medical director of a diabetes camp may order all injections of insulin be cut by 25% for a certain day because there will be a large amount of activity in the afternoon. This action by a medical director will likely result in the campers' blood glucose level remaining in a safe range, as activity and insulin both act to decrease blood glucose level. A specific general medical staff working with a specific camper may decide that camper should cut the dose of an injection by 25%, but cannot order such a change on a camp-wide scale. In table 1, D5 and D6 are examples of roles. Important subtypes of DCO:DIABETES CAMP MEDICAL STAFF ROLE include: diabetes camp medical director role, diabetes camp medical coordinator role, diabetes camp lead medical staff role, and diabetes camp general medical staff role. This then leads to the introduction of a fourth type of representational unit: the ones required to drive the logic of the application. Again, care must be taken, to the extent possible, that these units denote entities a realism-based ontology can commit to. In this case, it can be achieved by reference to entities of the type BFO:RELATIONAL QUALITY, such as the relational quality of authority between a person in whom an instance of diabetes camp medical director role inheres and other human beings in whom instances of other diabetes camp medical staff roles inhere. Analogously, there is a relational quality of authority between a person in whom an instance of diabetes camp medical coordinator role inheres and other human beings in which instances of other diabetes camp medical staff roles inhere, except for the person in whom an instance of diabetes camp medical role inheres. This allows to create an authority hierarchy among the diabetes camp medical staff.

Finding relevant mid-level ontologies. A major problem trainees are confronted with when attempting to define mid-level classes of their application ontology in terms of existing classes in other ontologies is the wide range of options, unfortunately mostly incompatible ones, which are suggested when searching through ontology repositories such as the BioPortal (<http://bioportal.bioontology.org/>) and Ontobee (<http://www.ontobee.org/>). Even more possibilities become available through literature searches.

Clearly diabetes camps in particular and camps in general are organized. Does that make such camp an 'organization'? Over 30 existing ontologies use that term for a variety of distinct entities, the most frequent ones being: (1) the process of organizing something, e.g. organizing AMIA 2017, (2) the result of such process, e.g. the conference AMIA 2017 itself, and (3) some collection of people (and for some ontologies also other entities such as buildings, statutory documents, ...), such as (3a) those involved in organizing in the first sense, e.g. the AMIA 2017 Organization Committee and (3b) everything involved once AMIA 2017 is ongoing such as its attendees, hosts, people involved in preparing the halls and rooms, etc. The student, in our use case, selected option (3a) which is the one at first sight suggested by the OBI, but with which the instructor, as we will explain further, disagreed (Table 1, D1a and D1b).

A similar situation occurs for 'blood glucose level' where the term is used to denote (1) an act of measuring glucose levels, (2) the value obtained through such measurement or (3) a quality of the blood independent of whether it is measured or not. In this case, the student correctly included all three types of entities in his analysis and definitions (Table 2, D9). Although several of the terms presented in the Clinical Measurement Ontology<sup>26</sup> may be of use to this work, including 'blood glucose level' itself, the OBI describes better the relations between the various entities. For example, the original OBI publication describes in great detail the process of collecting a sample of blood and analyzing the glucose content in the sample<sup>17</sup>. Thus, this ontology contains a class for OBI:MEASURING GLUCOSE CONCENTRATION IN SERUM so that the resulting output DCO:BLOOD GLUCOSE READING could be defined as a IAO:SCALAR MEASUREMENT DATUM which is linked to the former through the relationship *has\_specified\_output*.

Decisions on which ontologies to build further on should not only be taken on the basis of suitability of individual representational units for one's purpose, but on the quality of the ontology as a whole. Acceptance of the ontology in the OBO Foundry is often used as a quality argument but it needs to be taken with a pinch of salt. So was at some point the Vaccine Ontology<sup>30</sup> (VO) considered useful for the DCO. However, the VO considers the USA and Canada subtypes of 'country' rather than instances, which is too egregious a mistake for the quality label to be taken serious.

For the use case described, several ontologies with realism-based foundations are used. OMRSE is an ontology ‘cover[ing] the domain of social entities that are related to health care, such as demographic information and the roles of various individuals and organizations’<sup>18</sup>. Notably, subtypes of the OMRSE:HUMAN HEALTH CARE ROLE include OMRSE:PATIENT ROLE and OMRSE:HEALTH CARE PROVIDER ROLE as used in D5 and D6.

The OBI<sup>17</sup> turned out to be very useful for many definitions, for example to address ways in administering insulin. Historically, insulin is introduced into the body via a subcutaneous route. Most campers use insulin pumps, and a minority use insulin pens or syringes. Insulin pens are made of cartridges of insulin and a screw top upon which one can screw a needle to inject the insulin. Insulin pumps are small (pager-size and shape), battery-operated entities which can push insulin held in a syringe through a tube. The tube is inserted into the hypodermis of a person with a needle. A similarity between the two methods of insulin injection is that insulin pumps and insulin pens all *have as part* an OBI:SYRINGE. A key difference is that every instance of insulin pump *has part* some instance of OBI:COMPUTER. Since there are several sorts of insulin pumps, it is important for the diabetes camp medical staff to be aware of the manufacturer of a camper’s insulin pump, and what model it is, such that the medical staff will then be able to navigate the appropriate menu options in the envisioned application. For example, Rob Smith’s Medtronic Paradigm 751 would be an instance of DCO:MEDTRONIC INSULIN PUMP MODEL PARADIGM 751, which is a subtype of DCO:MEDTRONIC INSULIN PUMP, which is a type of DCO:INSULIN PUMP (Table 2, D10) which OBI:is *manufactured by value* Medtronic Minimed which itself is an *instance of* OBI:ORGANIZATION. Whether a DCO:CAMPER receives insulin via a DCO:INSULIN PUMP or DCO:INSULIN PEN, he or she is participating in the process OBI:SUBCUTANEOUS INJECTION.

The IAO makes it possible to keep entities on the side of the patient and caregivers, e.g. blood glucose levels, strictly separate from entities to describe such entities, e.g. blood glucose readings on the screen of a glucometer. So there is IAO:SETTING DATUM defined as ‘a datum that denotes some configuration of an instrument’. Insulin pumps are devices with programmable settings. Examples of settings which are programmed by a user include: insulin to carbohydrate ratio, correction factor (sensitivity), basal rates, and blood glucose level targets. The DCO includes these terms referring to the analogous setting datum as subtypes of a new term DCO:PUMP SETTING DATUM (Table 2, D10), which itself is a subtype of an IAO:SETTING DATUM. These setting datum provide the information to determine how much insulin should be given in a given situation, for example based on current blood glucose level of a camper and carbohydrates eaten by a camper. Other representational units based on IAO include DCO:Blood Glucose Reading (Table 2, D9) as a subtype of IAO:SCALAR MEASUREMENT DATUM.

Deciding what types of entities selected terms denote. This aspect of realism-based ontology development is undoubtedly the hardest nut for trainees to crack as it does not allow objects to be ‘anything perceivable or conceivable’ as advocated in concept-based approaches<sup>31, p845</sup>. So it would be tempting to create subtypes of OBI:SUBCUTANEOUS INJECTION for what terms such as ‘food bolus injection’ and ‘correction bolus injection’ would denote. A food bolus injection would be administered when a camper eats carbohydrates, and a correction bolus injection when a camper’s blood glucose is high. However, it is easy to conceive two processes of bolus injection which are of exactly the same type (same sort of needle, same concentration, same type of injection site, ...) whereby only the intention of the clinician would be different, but such intention is not part of the process of injecting. Or consider a carbohydrate count as subtype of IAO:SCALAR MEASUREMENT DATUM on the argument that such count is the result of a measurement of carbohydrates in food eaten. Carbohydrates are counted based on standard guidelines from the FDA and the nutrition labels on food. A standard may be: a medium size apple has 15 grams carbohydrates, and 1 slice of bread has 15 grams of carbohydrates. The question the trainee has to answer here is whether this method is more and *estimate* than a measurement, and whether OBI accepts estimates as kind of measurements.

As a last example, an important consideration when giving an injection of insulin is the location of the injection on the body. This is important to know, as different spots on the human body absorb insulin to varying degrees. For example, insulin injected into the abdomen is absorbed faster into the bloodstream than insulin injected into the buttocks, and a camper’s blood glucose level may be influenced by choosing to perform injections into certain locations. One could follow the example of the Vaccine Ontology<sup>30</sup> (VO), a domain ontology which represents the entities in reality as pertaining to vaccines and vaccinations. Since vaccines can be given in various locations in the human body and via various methodologies, for example, an injection of a vaccine into the deltoid muscle of the arm, or an inhalation of a vaccine through the nose, the VO defines a VO:VACCINATION SITE, and a subtype VO:VACCINE INJECTION SITE. Thus in a similar way DCO: INSULIN INJECTION SITE is defined (Table 2, D12, D13). The question to be answered here is whether such terms denote truly extra entities or whether such terms are mere ‘ways of talking’. After all: there is no injection site in BFO sense unless the injection has been given and one can therefore not use it to describe where an injection is *to be* given; the ‘site used for injection’ is simply a body part. It is a bit different for inhalation, because in that case, the body part (nostril, mouth, ...) is indeed a site in BFO sense. But would that site



become suddenly an instance of ‘insulin inhalation site’ after inhalation of insulin? What did change after the inhalation that warrants this? It is acceptable, for pragmatic reasons, to introduce what are called ‘defined classes’ for these purposes, but the resulting ontologies are then not ‘realism-based’ in the strict sense as defined by some authors<sup>32</sup>.

Closely related to this is determining whether entities are types or instances. Instances, also called ‘particulars’ or ‘individuals’, are entities that carry identity and do not come in various sorts. Canada and the USA are instances: there are no two different kinds of Canada or different kinds of USA. Is ‘Camp Aspire’, acronym for ‘*Always Sharing Priceless, Inspirational, Rewarding Experiences*’ as organized by the ADA a type or an instance (Table 1, D4)? Interestingly, that there is another ‘kind of’ ‘Camp Aspire’, acronym for ‘*Autism Social Skills Program for Interaction and Relationship Education*’ organized by the Autism Society of Central Illinois is not an argument for typehood but a simple case of homonymy. But that the ADA ‘Camp Aspire’ is organized annually, in distinct locations and at some locations at different times are all arguments for typehood, each specific camp, at a specific site and during a specific time period being an instance of that type. And of course, both the ADA ‘Camp Aspire’ and the autism ‘Camp Aspire’, which also is organized multiple times, are instances of DCO:CAMP. This raises the question whether ‘camp’ should indeed be defined in DCO, or whether it should not move up to a higher level in the hierarchy of domains.

What if the experts don’t know? To what ontology a representational unit ideally belongs is not a clear cut matter. It is proposed that the distinction between occurrents (entities that evolve, such as processes) and continuants (entities that don’t have temporal parts), and, amongst the latter, between independent continuants (e.g. objects) and dependent continuants (e.g. qualities, roles, functions, ...) should be a guide. The Gene Ontology has been used as an example of how that would work<sup>1, p129</sup>. But what if even the experts disagree – or can’t decide – about what the highest subsumer of some type might be. It was tempting to define DCO:CAMP (Table 1, D1) as a type of OBI:ORGANIZATION until the fine print was read: ‘*The definition summarizes long email discussions on the OBI developer, roles, biomaterial and denrie branches. It leaves open if an organization is a material entity or a dependent continuant, as no consensus was reached on that. The current placement as material is therefore temporary, in order to move forward with development.*’ ([http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI\\_0000245](http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI_0000245)). Also using OBI:HOMO SAPIENS to denote human beings is not unproblematic as experts seem to disagree about whether this OBO-term just lacks face validity by using the name of a specific species (an individual) for a type or truly represents a species<sup>33</sup>. The difference is important as there are no different kinds of homo sapiens (yet?).

## Conclusion

Building realism-based ontologies is a daunting task. Although the first version of BFO was created fifteen years ago, the current version consists of a mere 35 classes with a maximum hierarchy depth of 5. This is not only because BFO as top-level reference ontology is by design restricted to representational units which are relevant to all domains and thus is expected to be rather small, but also because of the extremely detailed philosophical analyses that were required to bring it into its present shape. Although not the same level of philosophical enquiry is needed for application ontologies, it cannot be avoided. The number of realism-based ontologies is growing enormously<sup>1, p160-162</sup>, but so also the number of inconsistent uses, even incompatibilities. Although ‘Building Ontologies with Basic Formal Ontology’<sup>1</sup> is a useful resource, it does not contain all the answers to practical questions trainees are confronted with.

## References

1. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. Cambridge, Massachusetts: The MIT Press,; 2015. Available from: ebrary <http://site.ebrary.com/gate.lib.buffalo.edu/lib/buffalo/Doc?id=11085695>.
2. Brochhausen M, Burgun A, Ceusters W, Hasman A, Leong TY, Musen M, et al. Discussion of "biomedical ontologies: toward scientific debate". *Methods Inf Med.* 2011;50(3):217-36.
3. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Appl Ontol.* 2010;5(3-4):139-88.
4. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology.* 2007;25(11):1251-5.
5. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine.* 1998;37(4-5):394-403.
6. IHTSDO. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English). 2015. p. 757.
7. Ceusters W, Elkin P, Smith B. Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach. *International Journal of Medical Informatics.* 2007;76:326-33.

8. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems. *Journal of the American Medical Informatics Association*. 2006;13(13):277-88.
9. Schulz S, Rector A, Rodrigues JM, Spackman K. Competing interpretations of disorder codes in SNOMED CT and ICD. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2012;2012:819-27.
10. Schulz S, Suntasirivaraporn B, Baader F. SNOMED CT's problem list: ontologists' and logicians' therapy suggestions. *Stud Health Technol Inform*. 2007;129(Pt 1):802-6.
11. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. *Summit on translational bioinformatics*. 2009;2009:116-20.
12. American Diabetes Association. Diabetes management at camps for children with diabetes. *Diabetes Care*. 2012;35 supplement 1:S72-S5.
13. Anderson B, Ho J, Brackett J, Finkelstein D, Laffel L. Parental involvement in diabetes management tasks: relationships to blood glucose monitoring adherence and metabolic control in young adolescents with insulin-dependent diabetes mellitus. *J Pediatr*. 1997;130(2):257-65.
14. Carlson KT, Carlson GW, Jr., Tolbert L, Demma LJ. Blood glucose levels in children with Type 1 diabetes attending a residential diabetes camp: a 2-year review. *Diabetic medicine : a journal of the British Diabetic Association*. 2013;30(3):e123-6.
15. Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. *Genome Biology*. 2005;6(5).
16. Smith B, Ceusters W. Aboutness: Towards Foundations for the Information Artifact Ontology. *International Conference on Biomedical Ontology*; July 27-30; Lisbon, Portugal 2015. p. 47-51.
17. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. *PLoS One*. 2016;11(4):e0154556.
18. Hicks A, Hanna J, Welch D, Brochhausen M, Hogan WR. The ontology of medically related social entities: recent developments. *J Biomed Semantics*. 2016;7:47.
19. Rahimi A, Liaw ST, Taggart J, Ray P, Yu H. Validating an ontology-based algorithm to identify patients with type 2 diabetes mellitus in electronic health records. *Int J Med Inform*. 2014;83(10):768-78.
20. Lin Y, Sakamoto N. Ontology driven modeling for the knowledge of genetic susceptibility to disease. *Kobe J Med Sci*. 2009;55(3):E53-66.
21. El-Sappagh S, Elmogy M, Riad AM. A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. *Artif Intell Med*. 2015;65(3):179-208.
22. Daghistani T, Shammari RA, Razzak MI. Discovering Diabetes Complications: an Ontology Based Model. *Acta Inform Med*. 2015;23(6):385-92.
23. Reddy SS. Evolving to Personalized Medicine for Type 2 Diabetes. *Endocrinol Metab Clin North Am*. 2016;45(4):1011-20.
24. Zhang Y-F, Tian Y, Zhou T-S, Araki K, Li J-S. Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems 2016.
25. Schweitzer M, Lasierra N, Hoerbst A. Observing health professionals' workflow patterns for diabetes care - First steps towards an ontology for EHR services. *Stud Health Technol Inform*. 2015;210:25-9.
26. Shimoyama M, Nigam R, McIntosh LS, Nagarajan R, Rice T, Rao DC, et al. Three ontologies to define phenotype measurement data. *Frontiers in genetics*. 2012;3:87.
27. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res*. 2015;43(Database issue):D1071-8.
28. El-Sappagh S, Ali F. DDO: a diabetes mellitus diagnosis ontology. *Applied Informatics*. 2016;3(1):5.
29. Rudnicki R, Ceusters W, Manzoor S, Smith B. What particulars are referred to in Electronic Health Record data? A case study in integrating Referent Tracking into an EHR application. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2007:630-4.
30. Lin Y, He Y. Ontology representation and analysis of vaccine formulation and administration and their effects on vaccine immune responses. *J Biomed Semantics*. 2012;3(1):17.
31. Shortliffe EH, Cimino JJ. *Biomedical informatics : computer applications in health care and biomedicine* 2014.
32. Smith B, Kusnierczyk W, Schober D, Ceusters W. Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. *KR-MED 2006, Biomedical Ontology in Action*. Baltimore MD, USA 2006.
33. Hogan WR, Ceusters W. Diagnosis, misdiagnosis, lucky guess, hearsay, and more: an ontological analysis. *J Biomed Semantics*. 2016;7(1):54.