

Original Paper

Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach

Xia Jing¹, MD, PhD; Nicholas R Hardiker², RN, PhD, FACMI; Stephen Kay³, PhD, FACMI; Yongsheng Gao⁴, MD, PhD

¹Department of Social and Public Health, College of Health Sciences and Professions, Ohio University, Athens, OH, United States

²School of Human and Health Sciences, University of Huddersfield, Huddersfield, United Kingdom

³THISC Ltd, Manchester, United Kingdom

⁴International Health Terminology Standards Development Organization, London, United Kingdom

Corresponding Author:

Xia Jing, MD, PhD

Department of Social and Public Health
College of Health Sciences and Professions
Ohio University

Grover Center W357

Athens, OH, 45701

United States

Phone: 1 7405930750

Fax: 1 7405930555

Email: xia.xjing@gmail.com

Abstract

Background: Ontologies are key enabling technologies for the Semantic Web. The Web Ontology Language (OWL) is a semantic markup language for publishing and sharing ontologies.

Objective: The supply of customizable, computable, and formally represented molecular genetics information and health information, via electronic health record (EHR) interfaces, can play a critical role in achieving precision medicine. In this study, we used cystic fibrosis as an example to build an Ontology-based Knowledge Base prototype on Cystic Fibrosis (OntoKBCF) to supply such information via an EHR prototype. In addition, we elaborate on the construction and representation principles, approaches, applications, and representation challenges that we faced in the construction of OntoKBCF. The principles and approaches can be referenced and applied in constructing other ontology-based domain knowledge bases.

Methods: First, we defined the scope of OntoKBCF according to possible clinical information needs about cystic fibrosis on both a molecular level and a clinical phenotype level. We then selected the knowledge sources to be represented in OntoKBCF. We utilized top-to-bottom content analysis and bottom-up construction to build OntoKBCF. Protégé-OWL was used to construct OntoKBCF. The construction principles included (1) to use existing basic terms as much as possible; (2) to use intersection and combination in representations; (3) to represent as many different types of facts as possible; and (4) to provide 2-5 examples for each type. Hermit 1.3.8.413 within Protégé-5.1.0 was used to check the consistency of OntoKBCF.

Results: OntoKBCF was constructed successfully, with the inclusion of 408 classes, 35 properties, and 113 equivalent classes. OntoKBCF includes both atomic concepts (such as amino acid) and complex concepts (such as “adolescent female cystic fibrosis patient”) and their descriptions. We demonstrated that OntoKBCF could make customizable molecular and health information available automatically and usable via an EHR prototype. The main challenges include the provision of a more comprehensive account of different patient groups as well as the representation of uncertain knowledge, ambiguous concepts, and negative statements and more complicated and detailed molecular mechanisms or pathway information about cystic fibrosis.

Conclusions: Although cystic fibrosis is just one example, based on the current structure of OntoKBCF, it should be relatively straightforward to extend the prototype to cover different topics. Moreover, the principles underpinning its development could be reused for building alternative human monogenetic diseases knowledge bases.

(*JMIR Med Inform* 2018;6(4):e52) doi: [10.2196/medinform.9979](https://doi.org/10.2196/medinform.9979)

KEYWORDS

cystic fibrosis; knowledge base; knowledge representation; molecular genetics information; ontology; OntoKBCF; phenotypes

Introduction

“The Semantic Web is a vision for the future of the Web, in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web” [1]. The major objectives of the Semantic Web are to improve data reuse, data sharing, and data integration in the data-centric Web era [2]. The World Wide Web, in contrast, is document-centric, whereby a document, not the data, is the basic processing unit. Therefore, the Semantic Web, owing to a finer granularity, has the potential to achieve a more precise result in its application.

Ontologies are key enabling technologies for the Semantic Web [3]. Ontologies play a key role in information use via knowledge sharing and reuse [4]. According to the Oxford Living Dictionary, an ontology is defined as “a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.” Another definition by Tom Gruber is “an ontology is a specification of conceptualization” [5]. In this paper, we use ontology to refer a collection of representations of the domain knowledge units (ie, entities and properties) and their relationships that can be utilized by humans, databases, and applications. The Web Ontology Language (OWL) is a semantic markup language for publishing and sharing data by ontologies on the Web [1,6]. OWL is intended to be used in information processing by machines, as well as understood by human beings [1,6]. Description logic can be subjected to automated reasoning to check the consistency of ontologies and determine subsumption relationships [6].

The specific application of ontologies includes standard conceptual vocabularies, services for queries, and reusable knowledge bases, all of which can facilitate interoperability across different systems [7]. Biomedicine is one of the active application domains of ontologies. The *motivation* of our work is to demonstrate the feasibility of providing formal and computable clinical information and molecular genetic information, derived from an ontology-based knowledge base, via an electronic health record (EHR) interface. Molecular genetic information plays a key role in medicine; this information, however, is not the kind that is traditionally included in an EHR system. Therefore, we aim to build an Ontology-based Knowledge Base prototype on Cystic Fibrosis (OntoKBCF) [8,9] to supply formal and machine-processable, clinically relevant molecular genetics information dynamically via an EHR prototype.

Cystic fibrosis is used as an example disease in building OntoKBCF for a feasibility demonstration. The reasons for this choice include the following: cystic fibrosis has a relatively stable and comprehensive molecular genetic profile; it is one of the most common, fatal genetic diseases in the United States [10], especially in white people of Northern European ancestry

[11,12]; the mechanism of the disease has been well studied and well understood; and cystic fibrosis is a single gene disease, which is less complicated than multiple-gene diseases. The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene causes cystic fibrosis.

An ontology-based knowledge base is built on top of domain entities, properties, and relationships (ie, the ontology); however, it also includes sharable instances, rules, and inference capabilities [13], which extend the ontology. Therefore, an ontology-based knowledge base can be utilized via an EHR system, for example, by incorporating the characteristics of patients. In this paper, we use a *knowledge base model* to refer to the logic specifications of a knowledge base at the abstract level and a *knowledge base prototype* (ie, OntoKBCF) to refer to the physical artifact that we eventually built at the concrete level, based on a knowledge base model. We use a *knowledge base prototype* instead of a *knowledge base* because we include different levels of facts in OntoKBCF, for example, nucleotide changes, amino acid changes, and clinical phenotypes. The facts are not exhaustive at any horizontal level within OntoKBCF; however, the facts are sufficient to organize the relationships at cross-horizontal levels [8,9] and to demonstrate how to generate a knowledge resource including molecular genetics information and health information via an EHR prototype [14,15].

An overview of the project, showing how OntoKBCF supplies computable and formally represented molecular and health information to an EHR in a dynamic manner has been published previously [14], as has the technical integration of OntoKBCF and the EHR prototype [15]. A further related source [8], published at an early stage of the project, mainly describes the *domain coverage* of OntoKBCF. In contrast, this paper provides *a holistic view of the construction principles* of OntoKBCF, including the *representations of the integration* between OntoKBCF and an EHR. As such, this paper may provide a reference for content developers operating in the Semantic Web space, especially in building knowledge bases and knowledge base modeling on monogenetic human diseases.

Methods

Representation Language and Terms Used in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

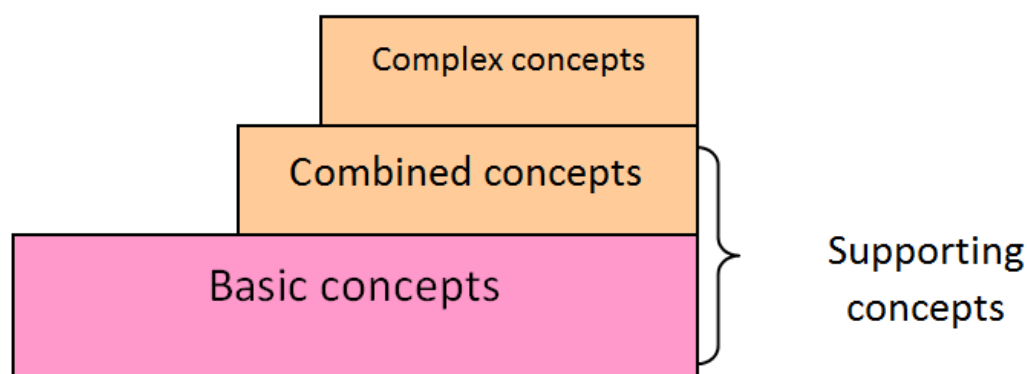
OntoKBCF was built using Protégé-OWL [16]. We chose OWL description logic (DL) [1] as the language for OntoKBCF. We used basic, combined, and complex concepts, as well as final facts, to represent selected content within OntoKBCF. **Table 1** provides definitions for the major terms that we have used in the remainder of the paper, with examples. **Figure 1** shows the relationships between the terms and their relative volumes of each category within OntoKBCF.

Table 1. Major construction terms used in the paper.

Term name	Annotation and example
Basic concept	The atomic concept from Unified Medical Language System [17], Gene Ontology [18], or domain knowledge, such as Gly, which is the amino acid glycine.
Combined concept	Combination of two or three basic concepts, such as Gly542, which is Gly's location, 542 in the cystic fibrosis transmembrane conductance regulator amino acid chain.
Complex concept	Combinations of more than three basic concepts and usually they are subjects. A subset of the combined concepts is explained through the EHR ^a interface, such as Patient_CF_with_Gly542X, which is a group of cystic fibrosis patients with the Gly542X variant.
Supporting concept	Basic concepts and combined concepts, used in explaining the complex concept.
Final fact	A domain statement represented in OntoKBCF ^b through a combination of basic concepts or combined concepts with properties and logic relationships, such as property and description of Patient_CF_with_Gly542X.
Fact	Any concept above is a subset of facts, which includes (1) hierarchy of concepts, both basic and combined concepts, and (2) a property description of the concept.

^aEHR: electronic health record.

^bOntoKBCF: Ontology-based Knowledge Base prototype on Cystic Fibrosis.

Figure 1. Relationships between construction terms used in this paper and their relative sizes.

Organization and Main Construction Procedures of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The main axes used in OntoKBCF are time and problem in line with the content of the included domain. The main axes are utilized to organize the atomic, combined, and complex concepts in OntoKBCF. Figure 2 shows the main construction processes for OntoKBCF.

Scope of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis and Sources

Figure 3 presents the main content of OntoKBCF. One of the purposes of OntoKBCF is for clinical use; therefore, the following factors were considered when determining domain content: (1) treatment or therapy, which has been reported as the main category of information needs for clinicians, including gene therapy and other treatments for cystic fibrosis [19-25]; (2) availability of relevant molecular genetic information, such as the most common *CFTR* mutations, particularly those related to health information or phenotypes (eg, symptoms, ethnic groups) [26-28]; (3) cystic fibrosis has a time-related development cycle, and usually, a patient's age is available in

an EHR; thus, time-oriented descriptions of cystic fibrosis are included in OntoKBCF [21,23-25].

Top-to-Bottom Content Analysis in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The selected content required analysis to set up the construction strategies. Based on the scope and the knowledge resources available, analyzing the facts by utilizing a series of dissections was necessary to make the task realistic and practically accomplishable. The starting point was sentences of facts that were dissected into atomic units (ie, basic concepts in Figure 1). The atomic units were represented with their necessary hierarchies, they were combined, and their properties included to express the complex facts in a logical, machine-processable format. Usually, complex facts can be represented by one or multiple concepts, described in Figure 1, and their properties.

The general criteria of the dissection were determined by the scope and granularity of OntoKBCF; otherwise, dissection would be an endless process. Only necessary superclass concepts, rather than detailed and complete subsets of Unified Medical Language System (UMLS) [17] and Gene Ontology (GO) [18], have been included in OntoKBCF. Only the properties and constraints within the scope of OntoKBCF, rather

than a complete and comprehensive description for each class, are represented. The main reason for doing so is to ensure that the tasks are accomplishable with limited resources and within

a restricted timeline. The dissection process was followed by reconstruction. Figure 4 shows the analysis and reconstruction order that we utilized in constructing OntoKBCF.

Figure 2. General construction procedures of the ontology-based knowledge base prototype on cystic fibrosis. KB: knowledge base; UMLS: Unified Medical Language System; GO: gene ontology.

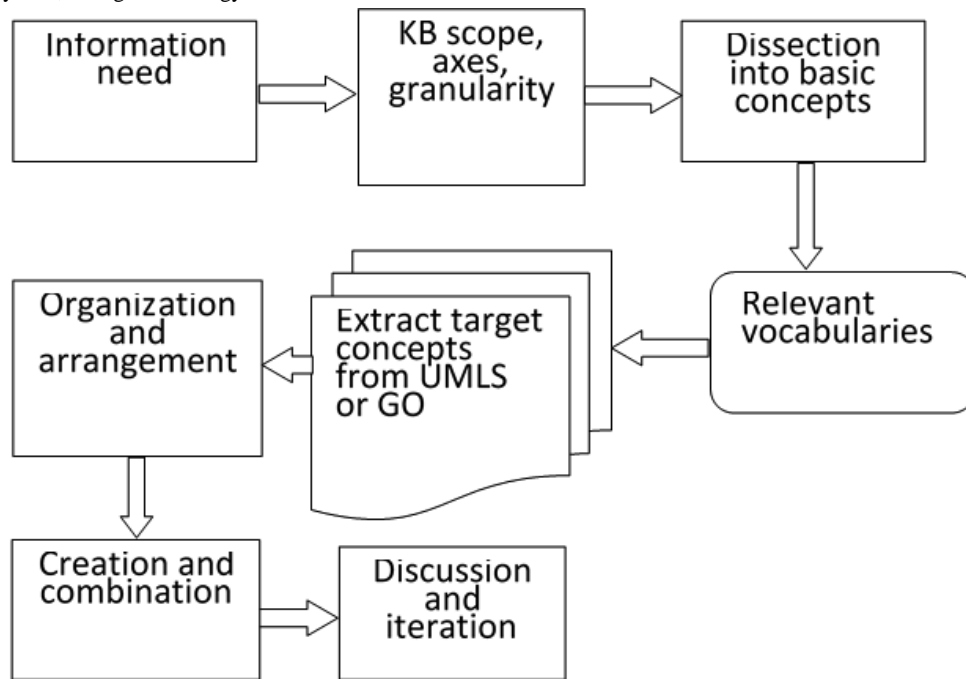
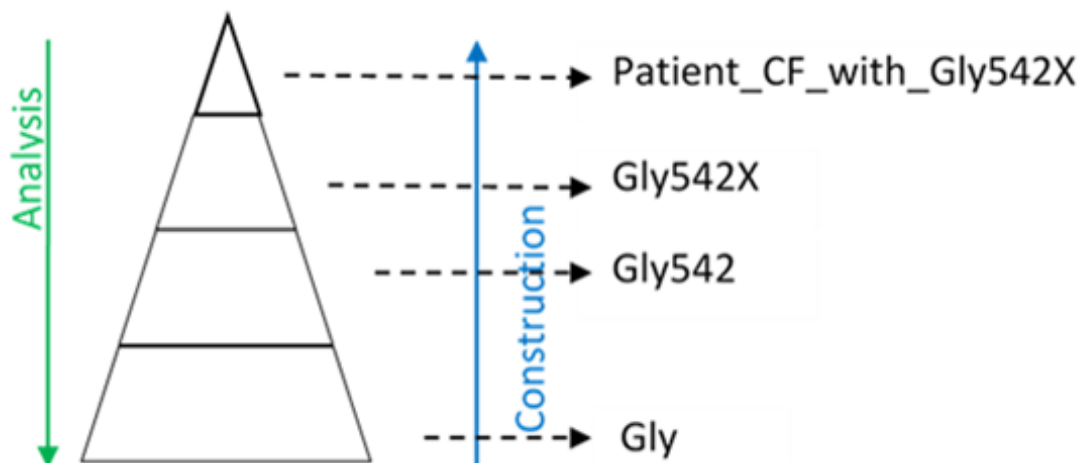


Figure 3. Main content categories represented in the ontology-based knowledge base prototype on cystic fibrosis and their potential usage.

	Patient characteristic	Potential usage by clinicians
Genotype-related	Mutation	Gene therapy
Phenotype-related	Age or sex	Treatment or Cochrane

Figure 4. General analysis (top-to-bottom) and construction (bottom-up) for ontology-based knowledge base prototype on cystic fibrosis.



Bottom-Up Construction Examples in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF was constructed bottom-up [29] as follows. After dissection analysis, reconstruction work started from basic atomic concepts. Then, the basic concepts were modified step-by-step until the basic concept was turned into a meaningful composite (a combined or complex concept). Every later concept used the former ones in its representation. Then, final facts were represented by the combination of a subject (ie, its definition) and its properties. A combination of basic concepts was utilized to build combined and complex concepts. The complex concept could be utilized as the subject to describe characteristics for a subgroup of patients. Figure 4 shows the construction process of OntoKBCF.

By way of example, if a group of patients with the Gly542X variant needs to be represented, the following steps would be used (see Figure 4):

- “Gly” stands for glycine, a type of amino acid; this is a basic concept.
- “X” is the nonsense codon, which terminates the translation; this is a basic concept.
- “Gly542” is defined as an amino acid location in the human *CFTR* amino acid chain; this is a combined concept.
- “Gly542X” is defined as an amino acid substitution in human *CFTR* amino acid chain; this is a combined concept.
- “Patient_CF_with_Gly542X” is defined as a group of cystic fibrosis patients with the amino acid change; this is a complex concept.
- Final facts about the group of cystic fibrosis patients with the Gly542X variant are represented by the combination of

both the logical definition of “Patient_CF_with_Gly542X” and its properties.

Most of the hierarchies of basic concepts in the knowledge base prototype follow UMLS and GO. In addition, relationships or classes were added or adjusted if there were no obvious choices. For example, in OntoKBCF, “sex_group” was used to connect “population_group” and “female,” and the concept “*CFTR* gene” in UMLS was adapted to “Human *CFTR* gene”; considering the clinical usage potential, we distinguished the human origins of the *CFTR* gene.

Intersection and Combination Representation Principles

We used the intersection among basic concepts to represent combined and complex concepts in OntoKBCF. For example, “female adolescent cystic fibrosis patient” is represented as a complex concept in OntoKBCF. Figure 5 shows the intersectional representation principle in OntoKBCF for female (ie, the purple circle), adolescent (ie, the rose circle), and cystic fibrosis patients (ie, the black circle) as a group, which represents *all* possible properties of this group. Figure 6 shows the exact representation (ie, final facts are represented as the combination of a definition of a complex concept and all its properties on the right lower corner of the figure) that we created within Protégé-OWL. The description in Figure 6 is a subset of the intersection in Figure 5. The entire set of properties for the subject (intersection of 3 bigger ellipses), compared with the filled ellipse (in blue) represented in OntoKBCF, refers to all the possible properties related to an adolescent (in rose) female (in purple) CF patient (in black). The reasons that we did not represent all possible properties are listed in the section Top-to-Bottom Content Analysis in OntoKBCF.

Figure 5. Representation of adolescent female cystic fibrosis (CF) patient in ontology-based knowledge base prototype on cystic fibrosis.

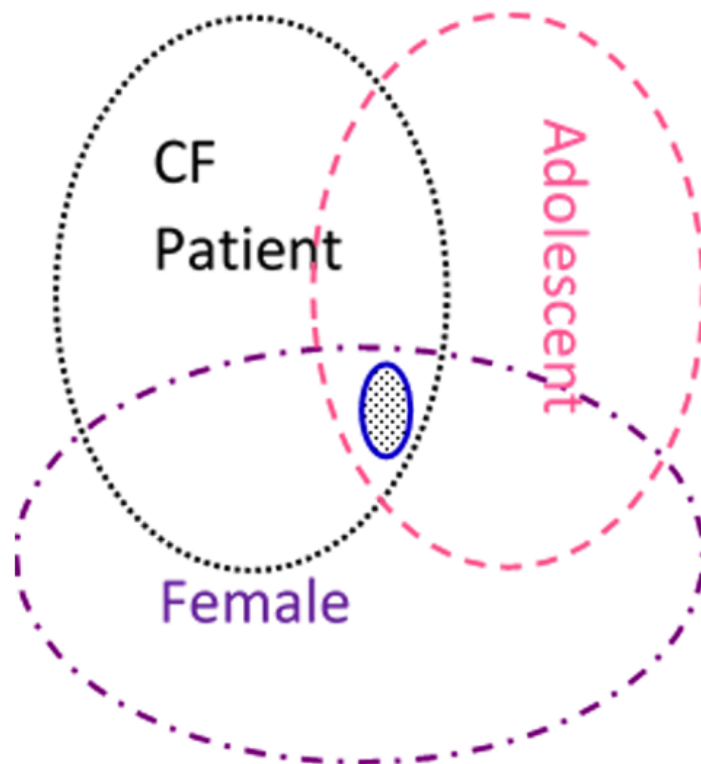


Figure 6. The representation of an adolescent female cystic fibrosis patient group in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Stanford University).

The screenshot shows the Protégé-Web Ontology Language interface. On the left is a class hierarchy tree with 'Adolescent_female_CF' selected. On the right, the details for 'Adolescent_female_CF' are shown, including an annotation: 'rdfs:comment [type: xsd:string] [infertile females with scanty cervical mucus]' and a description: 'Equivalent To Adolescent_CF and (occur_in_gender_group some Female)'. Other details include 'SubClass Of has_manifestation some (Infertility and Scanty_cervix_mucus)' and 'SubClass Of (Anonymous Ancestor) Human_being and (has_diagnosis some Cystic_fibrosis)'. The interface also shows tabs for 'Class Annotations' and 'Class Usage'.

Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Construction Principles

Our *philosophy* is to use the existing concepts as much as possible and to create new ones only if it is absolutely necessary. If a concept is from UMLS, then the unique concept identifier from UMLS is kept in the Annotations section for future

reference (eg, reuse, mapping, and communication). Other construction *principles* include the following: (1) to represent more types (eg, nucleotide deletion, insertion, transition, and transversion) of facts within the same level; (2) to represent 2-5 examples within the same type, for example, for the single-nucleotide transition of *CFTR*, we represented 5 examples in OntoKBCF instead of representing all existing facts

exhaustively; (3) to have the same representation listed in all possible hierarchies in OntoKBCF; for example, *AA2183_minus_G* is listed under *Del_A*, *Nucleotide_deletion_in_human_CFTR_gene*, *A_transition_G*, and *Nucleotide_transition_in_human_CFTR_gene*; and (4) to represent only the facts within the scope of OntoKBCF. The purpose of our work is to demonstrate the feasibility of the construction of a knowledge base by utilizing Semantic Web technology and demonstrating that the facts from the knowledge base can be utilized via an EHR interface dynamically. Therefore, representing the exhaustive list of variations of the same gene is not a priority. In addition, as most of the mutation types have been covered in the current OntoKBCF, we believe it to be extendable. To include an exhaustive list of the variations of the *CFTR* gene should be straightforward when following our construction principles and approaches.

Naming Conventions in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

To create the potential for broader usage, some of the classes were specifically named for human beings in OntoKBCF, such as “Human_CFTR_gene” and “Human_CFTR_gene_exon”; this differentiates them from any additional classes that may concern other organisms.

The Nomenclature for the Description of Sequence Variations [30] is followed in OntoKBCF. In addition, the following basic name rules are used in OntoKBCF: (1) for amino acid changes (ie, protein level), a 3-letter abbreviation name was used; (2) for coding DNA (cDNA)-level nucleotide changes, a 1-letter name (such as “A,” “T,” “C,” and “G”) was used. These design decisions are present in OntoKBCF as follows: (1) a mutation name, which strictly follows the nomenclature recommendation [30,31], has been kept in the “Annotation” section and (2) there were only cDNA- and protein-level descriptions of the *CFTR* mutation name, which did not include the RNA-level description. The reason is that cDNA is used to code protein, and RNA is more of a middle layer with regards to protein synthesis. Amino acids can be named in 1 or 3 letters. There are both amino acids and nucleotides in OntoKBCF, and 1-letter abbreviation names of amino acids can be confused with nucleotides. For example, A can be Alanine (an amino acid; the 3-letter abbreviation is Ala) or Adenine. Thus, we use 3-letter abbreviation names for amino acids and 1-letter names for nucleotides.

As there are 22 types of amino acids, there are many more possibilities for amino acid changes compared with nucleotide mutations; only the amino acid changes that fell within the scope of OntoKBCF were included. In contrast, because there are 8 types of nucleotides, which consist of DNA and RNA, with a smaller number of possible mutations, the complete set of possible nucleotide mutations in cDNA for deletion, insertion, transversion, and transition were included in OntoKBCF. The symbols “+” or “-” indicate the nucleotide position at the beginning of the intron (+) or the end of the intron (-); this would be represented as “plus” and “minus” in OntoKBCF for nucleotide substitutions. Amino acid substitutions are handled similarly.

In OntoKBCF, the final expression of nucleotide mutations can be distinguished from amino acid changes as follows: (1) their labels contained “minus” or “plus,” or “Ins” or “Del” (such as “AA2183_minus” or “G621_plus_1”, and “Del394” or “Ins3905”); or (2) a 3-letter abbreviation name was used for amino acids to avoid confusion. For example, Gly was utilized as the abbreviation for glycine, not G, which can be confused with the nucleotide “G.” “Minus” and “plus” were used to represent the nucleotide position. “Ins” is used for nucleotide insertion, and “Del” is used for nucleotide deletion. In theory, not all nucleotide mutations have to be labeled with one of these 4 values; however, no such example has been found in the construction of OntoKBCF. This strategy avoids the potential limitations of the special characters that one can use in a name.

Other Considerations in Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

For properties in the OWL representation, only existential restriction “someValuesFrom” was used as the property restriction (appearing as “some” in the text); it would be inappropriate to use allValuesFrom (appearing as “only” in the text,) because it cannot be guaranteed that their value classes are fully specified and mutually exclusive and exhaustive. It is sufficient to use existential restriction in most representations to leave space for future efforts to complete the description. For example, for a cystic fibrosis patient with Gly542X, the subject had the mutation property “some Gly542X” because it is possible that Gly542X is not the only mutation for this type of patient.

According to Horridge et al [32], the “Domain” and “Range” conditions in the “Property” tab of the Protégé-OWL interface are not used as a constraint. These properties are outside of the scope of this project, and consequently, the “Domain” and “Range” widgets were left blank for all properties.

Results

Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Content Summary

Table 2 summarizes the metrics of OntoKBCF. Description logic expressivity was ALCHIF (attribute language; complex concept negation; role hierarchy; inverse properties; functional properties).

In OntoKBCF, the most fine-grain level of biological information starts at nucleobase, which is the most important component for the elementary units (nucleotides) of RNA and DNA. In contrast, for health information in OntoKBCF, the most fine-grain level starts from relatively *atomic concepts*, such as diarrhea, nausea, or coughing.

Currently, OntoKBCF includes answers to only questions about “what” (eg, what is Gly542X? What is the hierarchy of cystic fibrosis?), not about “how” (eg, disease mechanisms and disease pathways) or “why” (eg, reasons and explanations) [8]. This is a pragmatic approach to make this study manageable and feasible. We focus on “what” questions initially. We do recognize the importance of “how” and “why” questions in clinical usage, especially in explanation of pathophysiological

mechanisms and disease pathways. However, we leave these questions for future development. [Textbox 1](#) lists the main categories of content in OntoKBCF. In the [Multimedia Appendix 1](#), we also include a screenshot of all the object properties in OntoKBCF.

OntoKBCF has been shared via BioPortal [9]. We used a reasoner, Hermit 1.3.8.413 within Protégé-5.1.0, to check the consistency of OntoKBCF.

Ontology-Based Knowledge Base Prototype on Cystic Fibrosis Application Functions in an Electronic Health Record

OntoKBCF has been demonstrated to be able to supply personalized, nonduplicate molecular genetic information and health information according to a patient's characteristics

successfully, dynamically, and automatically via an EHR prototype [14,15]. The connection between OntoKBCF and the EHR prototype can be made through a combination of automatic and manual means. OntoKBCF includes a section for the EHR main structure, which is utilized to map the domain knowledge in OntoKBCF to the EHR structure and to make the connection automatic.

OntoKBCF can provide reference facts (eg, hierarchies of basic concepts, combined concepts, and final facts) and reminders to EHR users. When a user applies the final facts via EHR interfaces, all the basic concepts and combined concepts utilized for building a complex concept can be accessed. [Figure 7](#) shows a conceptual architecture about how OntoKBCF can be utilized within an EHR user interface.

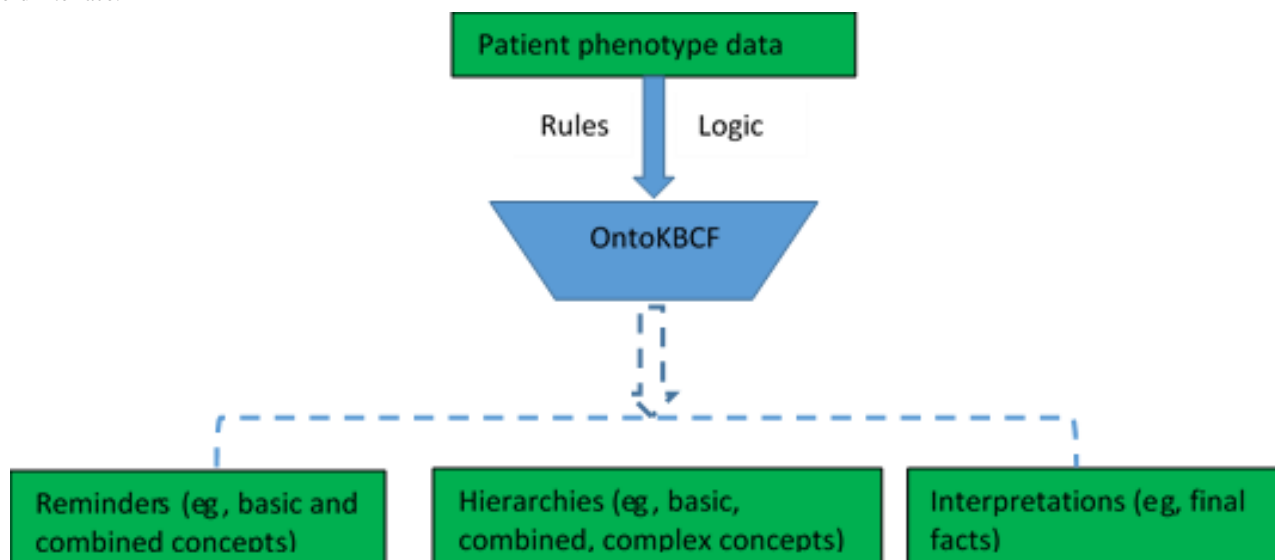
Table 2. Metrics of ontology-based knowledge base prototype on cystic fibrosis content summary [9].

Name	Count (n)
Axiom	1375
Logical axiom	665
Declaration axiom	445
Class	408
Individual	0
Object property	35
SubClassOf	498
EquivalentClasses	113
DisjointClasses	37
Maximum depth	9
Maximum number of children	20
Average number of children	3
Classes with a single child	69
AnnotationAssertion	265

Textbox 1. Main categories of content in ontology-based knowledge base prototype on cystic fibrosis.

<p>Biological information (genotype-related)</p> <ul style="list-style-type: none"> • Nucleotide mutation • Amino acid change • Gene mutation location • Gene or protein mutation dissection <p>Health information (phenotype-related)</p> <ul style="list-style-type: none"> • Demographic data (eg, age, sex, and ethnicity) • Symptoms • Diagnostic tests or diagnosis • Treatment facts <p>Electronic health record data representation structure</p>

Figure 7. The conceptual architecture between ontology-based knowledge base prototype on cystic fibrosis (OntoKBCF) and an electronic health record interface.



Discussion

Significance of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF meets our design objectives and has been integrated successfully with an EHR prototype [15]. This paper illustrates the construction principles, representation approaches, and challenges in constructing OntoKBCF in Protégé-OWL. We believe that the methods we reported to construct OntoKBCF will provide a useful reference for Semantic Web content developers, particularly in constructing knowledge bases on human monogenetic diseases for clinical usage. The basic and combined concepts and similar facts in both biological and health fields represented in OntoKBCF can be found in other diseases. OWL statements make OntoKBCF sharable and computer processable. The possibility of automated reasoning assures the consistency of the underlying ontology. The organization, structure, and use of existing, controlled vocabularies and terminologies as much as possible within OntoKBCF improve its potential for compatibility and reusability.

Patient Group Representation Challenge Example

In OntoKBCF, different groups of cystic fibrosis patients are represented separately. For example, there are cystic fibrosis patients in different age groups or with different genetic variations. Figure 8 shows the main categories of cystic fibrosis patient groups included in OntoKBCF.

Currently, for each cystic fibrosis patient group, we represent only the available facts from the selected resources and within the scope of OntoKBCF. We defined patient groups logically, rather than representing them comprehensively. Some cystic fibrosis patient groups, such as an adolescent cystic fibrosis patient group (Figure 9), are fairly straightforward, while some are fairly complicated because of available facts. Figure 6

demonstrates a partial representation of the adolescent female cystic fibrosis patient group.

For the complicated groups in OntoKBCF, the representation is far from comprehensive or exhaustive. We feel that if we were to set comprehensive representation as a goal, the task would be endless, as there are many properties related to the patient groups. The properties can include, among other things, clinical manifestations, treatment plans, nursing care plans, patient self-care plans and patient education. The detailed categories, as well as the actual representation for each property, would result in an exhaustive list. Our pragmatic solution to this scalability problem is to focus on practical applications, stay within the defined scope, and make it possible to expand in the future.

Ambiguous Concepts and Uncertain and Negative Statement Representation Challenge

There are some common expressions in clinical medicine, for example, “sometimes,” “especially,” “very common,” and “very rarely” that are challenging to represent in a computable and formal manner as their interrelationships are vague. These expressions are very important for human experts (such as physicians) to absorb or to apply the information. For example, for older children with cystic fibrosis, one important symptom is “nasal polyps, especially if recurrent” [21]. However, we cannot represent this type of fact in a computable form. Likewise, Tao et al [33] recognized that representing uncertain expressions is a challenge in creating a temporal-related ontology.

In addition to uncertain concepts, negative descriptions cannot be represented precisely, either. Considering the limited reasoning support, less than ideal reasoning efficiency, and the complexity in modeling concepts, currently, for demonstration purposes, we included only the definite and positive results in OntoKBCF. We do realize that although this is necessary, it is a limitation of OntoKBCF as it does not reflect the real world comprehensively.

Figure 8. Main cystic fibrosis patient groups in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Stanford University).

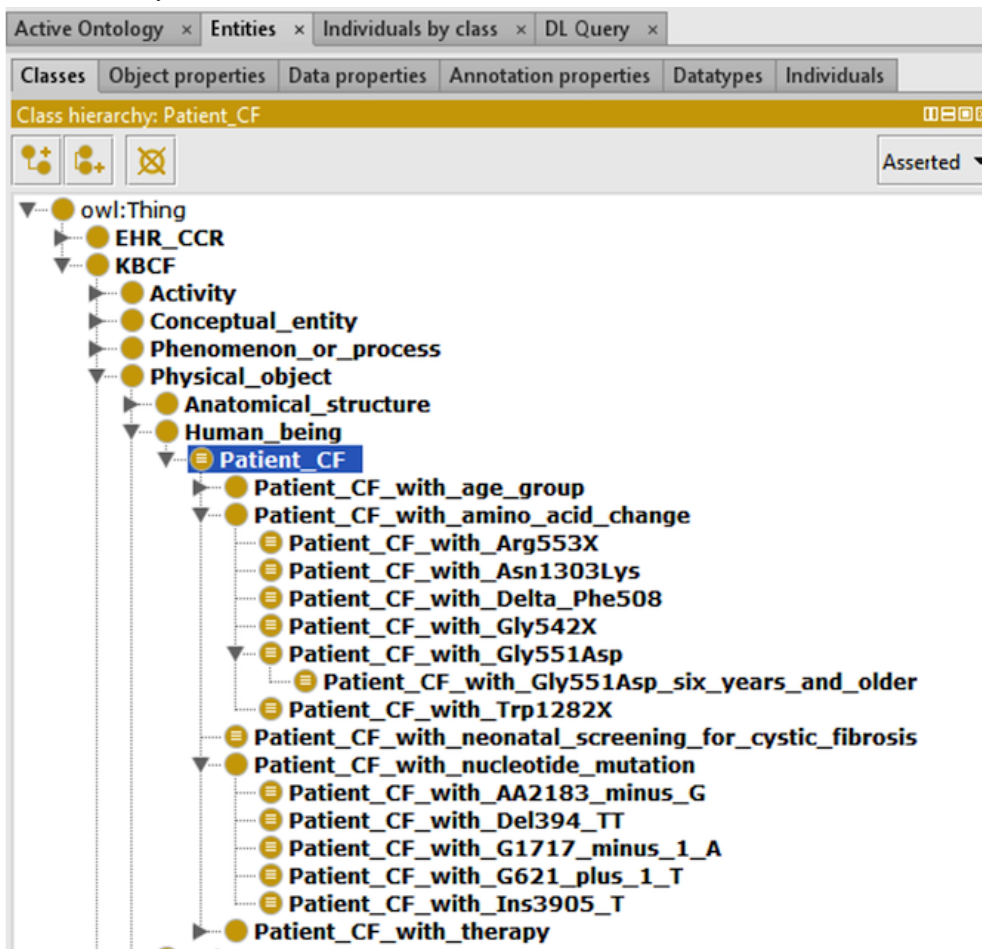


Figure 9. The representation of an adolescent cystic fibrosis patient group in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Stanford University).

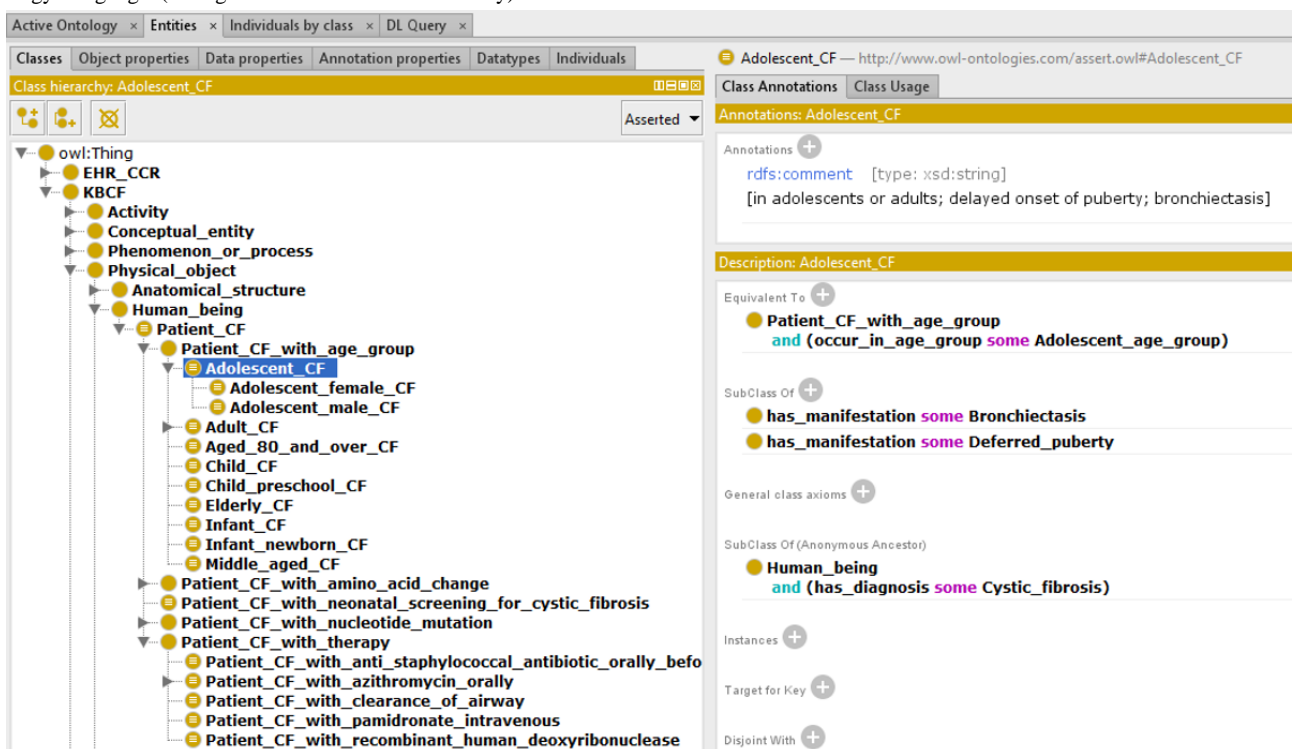
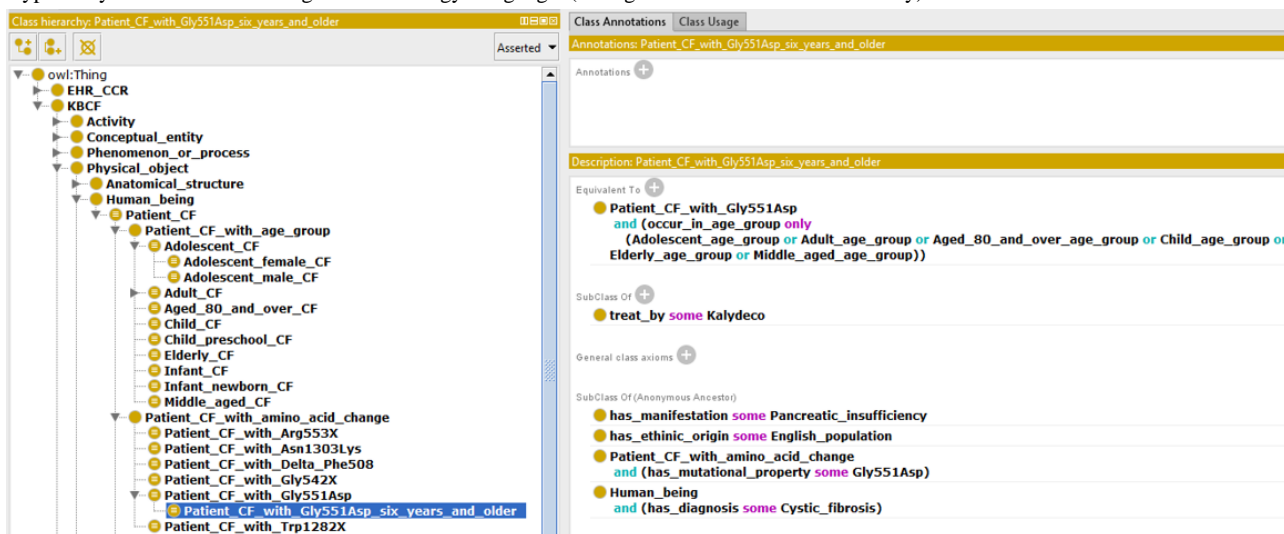


Figure 10. A cystic fibrosis patient group with the Gly551Asp variation and aged at least 6 years represented in ontology-based knowledge base prototype on cystic fibrosis via Protégé-Web Ontology Language. (Protégé Source: Stanford University).



Complex Facts Representation Challenges

One of the complex facts represented in OntoKBCF is shown in Figure 10. This is a representation of a cystic fibrosis patient group whose patients have a mutation from Gly to Asp at 552 positions and are at least 6 years old. The current properties include diagnosis, treatment (ie, medication) information, age groups, ethnic origin, clinical manifestation, and exact amino acid change. This example has positive and definite statements, which are represented in the current OntoKBCF. However, how to represent more complicated facts, such as detailed molecular mechanisms or pathway information, with higher than first-order logic and how to process those descriptions also would be challenging.

Broader Challenges in the Construction of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

The challenges of introducing information technologies and computer science applications to medicine have been recognized for a long time [34]. Knowledge representation in medicine is one of the many areas that face issues. It is difficult to say that the challenge is in knowledge representation, *per se*, or that the challenges are in the broader area of medicine, logic, or even mathematics.

Regarding OntoKBCF, we sought to determine how many properties are sufficient or ideal to describe a cystic fibrosis patient group. From a pragmatic point of view, setting up the scope of the project is necessary to make the project manageable. Otherwise, the representation would become an endless task, as there are many possible properties related to a group of cystic fibrosis patients, as Davis et al [35] described. If we assume that we constructed OntoKBCF and others constructed similar knowledge bases about cystic fibrosis, then other management-related challenges would emerge, for example, (1) how to validate classes from different sources; (2) how to match the same classes from different sources; (3) how to integrate the properties of the same class that originates from different sources; (4) how to integrate different ontologies about the same topic; and (5) how to authorize and achieve consensus

for the combined classes or ontologies conveniently. Currently, BioPortal [36] is an effective way to collect submitted biomedical ontologies; it can be utilized to check existing ontologies before building any new ontologies. Ontobee [37] provides a finer search on ontology terms, which goes further than index-based searches on ontologies, by providing an accessible unique resource identifier (URI) for ontology terms. These examples provide great potential in reusing and mapping ontology terms; however, many of the abovementioned challenges are still not solved. In addition to using URI, including the coded concepts, such as unique identifiers in UMLS or the clinical terminology systematized nomenclature of medicine-clinical terms, may improve the reusability of the knowledge base in general.

One critical challenge in building OntoKBCF is a lack of sufficient clinical actionable knowledge that is related to molecular genetic information. This is one of the main reasons that we constructed OntoKBCF as a prototype rather than as a product knowledge base.

Other challenges that we faced during the construction of OntoKBCF included the following: (1) how to communicate the construction principles and design considerations in a consistent, explicit, and sharable manner; (2) whether we can make the construction process automatic; and (3) how to update the domain content automatically. Solving these challenges may help us build larger-scale ontologies and help in their integration automatically.

Other Related Work

Our work was originally conceived in 2004 and mostly completed by 2009. At the time, there were no available resources that could meet *all* of the following criteria: consistent molecular genetic information and a clinical actionable, machine-processable format that was shareable, reusable, and customizable. We, thus, had to create our own resource to meet all the requirements. Biobanks might be a good source to provide such information. Nevertheless, the UK Biobank [38] started to recruit participants only in 2006, and the USA Biobank, as

a part of the Precision Medicine Initiative, was founded only in 2016 [39].

We used an ontology-based knowledge base to provide customizable molecular genetics and health information to EHR settings successfully. The ontology-based knowledge base provides potential in reusing and sharing, as well as the consistency of information. The creation of knowledge resources with fine granularity (not only disease names) has been recognized repeatedly [40,41] as the main challenge in bringing new information (eg, molecular genetics information) to an EHR. Although there are many existing databases for both genotypes and phenotypes of human beings, not all of them are organized in a machine-processable format. Therefore, the detailed construction principles and approaches for OntoKBCF reported in this paper can provide a reference to peers in the field.

One recent example of knowledge base was reported by Samwald et al [42,43] who built a resource description framework or OWL knowledge base to provide support for clinical pharmacogenetics. Their work shares some similarities with OntoKBCF. For example, it provides consistent genomics information in clinical settings in a machine-processable format. Both projects use reasoners to maintain consistency. The two projects, however, have different focused application domain areas; Samwald et al's work focuses on pharmacogenomics information, and OntoKBCF uses cystic fibrosis to demonstrate possible broader applications. In addition, Samwald et al's work includes a query and answer part that supersedes OntoKBCF. Meanwhile, OntoKBCF demonstrates its usage via an EHR prototype; such demonstration was not included in Samwald et al's [42,43] publications. There is no detailed description of the knowledge base construction in Samwald et al's project. Thus, it is difficult to compare the representation and construction principles and approaches of the two projects in depth.

In recent years, the broad importance of Semantic Web technologies has been recognized. There are many more studies that have used ontology-based knowledge bases to assist clinical tasks within an EHR. For example, Robles-Bykbaev et al [40] reported the use of a formal knowledge base model to assist in the generation of decision making and recommendations for communication disorders. They stated that their knowledge base is used to support data analysis and inference processes; however, the paper does not include the details of the organization and structure of the knowledge base. The Clinical Narrative Temporal Relation Ontology (CNTRO) [33,44] is a temporal-related ontology in the OWL. CNTRO is used for inference purposes in processing clinical narratives. The use cases [45] show promising results in processing short and simple adverse event narratives. The OWL query application [46] for

CNTRO and the harmonization of CNTRO with other existing time ontologies may improve CNTRO [44] further and provide broader applications in analyzing clinical narratives. Another example is the research of Wang et al [41] and Hu et al [47]. They utilized an ontology-based clinical pathways knowledge base to generate personalized clinical pathways for clinicians by incorporating patients' data. The clinical pathways knowledge base is independent of an EHR, and it can be shared by other systems. All these examples used ontology-based knowledge bases and integrated them with an EHR. There are other ontology examples that are utilized in settings other than in EHR, such as in mobile devices [48] or to facilitate social media data mining [49] for health purposes. While these recent studies demonstrate the general application of semantic technologies in health care, they do not consider the important clinical usage of relevant molecular genetic information.

Limitations of Ontology-Based Knowledge Base Prototype on Cystic Fibrosis

OntoKBCF is a knowledge base prototype rather than a knowledge base, which may include substantial instances for each class. OntoKBCF, however, includes content from nucleotides to clinical phenotypes across different granularity levels. If we can image a pyramid that moves from gene to protein to cell to tissue to organ up to human beings [50], we can see that the current OntoKBCF is not thorough at any horizontal level; however, it includes all vertical levels of knowledge. The expansion of each level should not be too challenging, as each horizontal level has some classes represented in OntoKBCF.

There has been no formal evaluation of OntoKBCF. However, we do have two limited levels of validation of the work: (1) automated reasoning assures the consistency of underlying ontology and (2) successful demonstration that OntoKBCF can deliver customizable molecular information and health information dynamically via an EHR prototype. As a knowledge base model, these levels of validation provide evidence that the original design objectives of OntoKBCF were met successfully.

Conclusion

We have presented the robust construction principles, approaches, design decisions, and challenges in constructing OntoKBCF. OntoKBCF is constructed in OWL and can be utilized as a knowledge resource to provide information on molecular genetics and other health topics dynamically to an EHR. The construction principles and approaches can be referenced for other topic areas, and based on its current structure, OntoKBCF could be expanded in a straightforward manner to a full knowledge base.

Acknowledgments

The authors would like to thank Mr Tom Marley for his contribution to the project and Dr Miao Sun for her patience, support, and expert knowledge in molecular genetics. Sincere thanks go to Prof James J Cimino for highly constructive discussions and valuable suggestions offered for the project. The work was funded by the University of Salford and partially through Ohio University, College of Health Sciences and Professions' new faculty startup funds. This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional screenshot of properties of OntoKBCF.

[\[PNG File, 310KB-Multimedia Appendix 1\]](#)

References

1. McGuinness DL, van Harmelen F. World Wide Web Consortium (W3C). 2004 Feb 10. OWL Web Ontology Language Overview URL: <https://www.w3.org/TR/owl-features/> [accessed 2006-03-30] [WebCite Cache ID 5EmVp1112]
2. Updegrove C. ConsortiumInfo.org. 2005 Jun. The Semantic Web: An Interview With Tim Berners-Lee URL: <http://www.consortiuminfo.org/bulletins/semanticweb.php> [accessed 2018-01-29] [WebCite Cache ID 6wq92rzvh]
3. Fensel D, Davies J, Harmelen F. Towards the Semantic Web: Ontology-driven Knowledge Management. Hoboken, NJ: Wiley, John & Sons, Incorporated; 2002.
4. World Wide Web Consortium (W3C). World Wide Web Consortium Issues RDF and OWL Recommendations: Semantic Web emerges as commercial-grade infrastructure for sharing data on the Web URL: <https://www.w3.org/2004/01/sws-pressrelease> [accessed 2018-01-29] [WebCite Cache ID 6wq9PIUMm]
5. Gruber T. Stanford Knowledge Systems, AI Laboratory. 1999. What is an Ontology? URL: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> [accessed 2009-03-08] [WebCite Cache ID 5f87rn7k1]
6. W3C OWL Working Group. World Wide Web Consortium (W3C). 2012 Dec 11. OWL 2 Web Ontology Language Document Overview (Second Edition) URL: <https://www.w3.org/TR/owl2-overview/> [accessed 2012-03-29] [WebCite Cache ID 66VqVQDIO]
7. Ozsu MT, Liu L. Encyclopedia Of Database Systems. New York City, NY: Springer-Verlag New York Inc; 2018.
8. Jing X, Kay S, Hardiker N, Marley T. Ontology-based knowledge base model construction-OntoKBCF. Stud Health Technol Inform 2007;129(Pt 1):785-790. [Medline: 17911824]
9. Jing X, Kay S, Hardiker N, Marley T, Cimino JJ. NCBO BioPortal. 2013 Jan 16. Ontological Knowledge Base Model for Cystic Fibrosis --BioPortal Internet URL: <http://bioportal.bioontology.org/ontologies/ONTOKBCF> [accessed 2018-06-29] [WebCite Cache ID 70XneQXIb]
10. National Human Genome Research Institute. Learning About Cystic Fibrosis URL: <https://www.genome.gov/10001213/learning-about-cystic-fibrosis/> [accessed 2018-01-29] [WebCite Cache ID 6wq9nLDB7]
11. Mayo Clinic. Cystic fibrosis URL: <https://www.mayoclinic.org/diseases-conditions/cystic-fibrosis/symptoms-causes/syc-20353700> [accessed 2018-01-29] [WebCite Cache ID 6wqAnQW0e]
12. Genetics Home Reference: NIH. Cystic fibrosis URL: <https://ghr.nlm.nih.gov/condition/cystic-fibrosis> [accessed 2018-01-29] [WebCite Cache ID 6wqBiqmu]
13. Bassiliades N, Vlahavas I. Active Knowledge-Based Systems: Techniques and Applications. In: Leondes CT, editor. Knowledge Engineering: Systems Techniques and Applications. London: Academic Press; 2000.
14. Jing X, Kay S, Marley T, Hardiker NR, Cimino JJ. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an EHR prototype based on the Continuity of Care Record standard. J Biomed Inform 2012 Feb;45(1):82-92 [FREE Full text] [doi: 10.1016/j.jbi.2011.09.001] [Medline: 21946299]
15. Jing X, Kay S, Marley T, Hardiker NR. Integration of an OWL-DL knowledge base with an EHR prototype and providing customized information. J Med Syst 2014 Sep;38(9):75. [doi: 10.1007/s10916-014-0075-4] [Medline: 24997857]
16. Musen MA, Protégé T. The Protégé Project: A Look Back and a Look Forward. AI Matters 2015 Jun;1(4):4-12 [FREE Full text] [doi: 10.1145/2757001.2757003] [Medline: 27239556]
17. Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-291 [FREE Full text]
18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000 May;25(1):25-29 [FREE Full text] [doi: 10.1038/75556] [Medline: 10802651]
19. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007;24(2):78-94.
20. Arroll B, Pandit S, Kerins D, Tracey J, Kerse N. Use of information sources among New Zealand family physicians with high access to computers. J Fam Pr 2002;51(8):706.
21. Harris A, Super M. Cystic fibrosis: the facts. 3rd ed. Oxford: Oxford University Press; 1995.
22. Orenstein D, Rosenstein B, Stern R. Cystic fibrosis: medical care. Philadelphia: Lippincott Williams & Wilkins; 2000.
23. Cystic Fibrosis Foundation. Adding Tomorrows and Living Today URL: <https://www.cff.org/> [accessed 2006-04-07] [WebCite Cache ID 5Ey9oAgZ9]

24. Cystic Fibrosis Trust. What is CF Internet URL: <https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis> [accessed 2018-01-29] [WebCite Cache ID 6wqC9e8bP]
25. Rosenstein BJ. The Merck Manuals. Cystic Fibrosis URL: <http://www.merckmanuals.com/professional/pediatrics/cystic-fibrosis-cf/cystic-fibrosis> [accessed 2018-01-29] [WebCite Cache ID 6wqCTk96o]
26. Zielenski J. Genotype and phenotype in cystic fibrosis. *Respiration* 2000;67(2):117-133. [doi: [10.1159/000029497](https://doi.org/10.1159/000029497)] [Medline: [10773783](https://pubmed.ncbi.nlm.nih.gov/10773783/)]
27. Zielenski J, Tsui L. Cystic fibrosis: genotypic and phenotypic variations. *Annu Rev Genet* 1995;29:777-807. [doi: [10.1146/annurev.ge.29.120195.004021](https://doi.org/10.1146/annurev.ge.29.120195.004021)] [Medline: [8825494](https://pubmed.ncbi.nlm.nih.gov/8825494/)]
28. Cystic Fibrosis: Mutation Database. URL: <http://www.genet.sickkids.on.ca/cftr/> [accessed 2005-12-07] [WebCite Cache ID 5C0czfzNe]
29. Rector A. School of Computer Science: The University of Manchester. 2005. Foundations of the Semantic Web: Ontology Engineering Internet URL: <http://www.cs.man.ac.uk/~rector/modules/CS646-2005/> [accessed 2018-01-29] [WebCite Cache ID 6wqCrJkUB]
30. den Dunnen JT, Antonarakis S. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;15(1):7-12. [doi: [10.1002/\(SICI\)1098-1004\(200001\)15:1<7::AID-HUMU4>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1098-1004(200001)15:1<7::AID-HUMU4>3.0.CO;2-N)] [Medline: [10612815](https://pubmed.ncbi.nlm.nih.gov/10612815/)]
31. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* 2016 Dec;37(6):564-569. [doi: [10.1002/humu.22981](https://doi.org/10.1002/humu.22981)] [Medline: [26931183](https://pubmed.ncbi.nlm.nih.gov/26931183/)]
32. Horridge M, Knublauch H, Rector A, Stevens R, Wroe C. CO-ODE Project. A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1 URL: http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf [accessed 2009-03-08] [WebCite Cache ID 5f88oV7Wk]
33. Tao C, Wei W, Solbrig H, Savova G, Chute C. CNTR0: a semantic web ontology for temporal relation inferencing in clinical narratives. In: *AMIA Annu Symp Proc. 2010 Presented at: AMIA 2010; Nov 13-17; Washington DC p. 787.*
34. Shortliffe E, Cimino J, editors. *Biomedical informatics: computer applications in health care and biomedicine*. 4th ed. London: Springer; 2014.
35. Davis R, Shrobe H, Szolovits P. What is a knowledge representation? *AI Magazine* 1993;14(1):33. [doi: [10.1609/aimag.v14i1.1029](https://doi.org/10.1609/aimag.v14i1.1029)]
36. Whetzel P, Noy N, Shah N, Alexander P, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res* 2011 Jul;39(Web Server issue):W541-W545 [FREE Full text] [doi: [10.1093/nar/gkr469](https://doi.org/10.1093/nar/gkr469)] [Medline: [21672956](https://pubmed.ncbi.nlm.nih.gov/21672956/)]
37. Ong E, Xiang Z, Zhao B, Liu Y, Lin Y, Zheng J, et al. Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Res* 2017 Jan 04;45(D1):D347-D352 [FREE Full text] [doi: [10.1093/nar/gkw918](https://doi.org/10.1093/nar/gkw918)] [Medline: [27733503](https://pubmed.ncbi.nlm.nih.gov/27733503/)]
38. UK Biobank. URL: <http://www.ukbiobank.ac.uk/> [accessed 2018-06-29] [WebCite Cache ID 70XnSxIey]
39. NIH Office of the Director. NIH funds biobank to support Precision Medicine Initiative Cohort Program Internet URL: <https://www.nih.gov/news-events/news-releases/nih-funds-biobank-support-precision-medicine-initiative-cohort-program> [accessed 2018-01-29] [WebCite Cache ID 6wqDkX0dV]
40. Robles-Bykbaev V, López-Nores M, Pazos-Arias J, Quisi-Peralta D, García-Duque J. An Ecosystem of Intelligent ICT Tools for Speech-Language Therapy Based on a Formal Knowledge Model. *Stud Health Technol Inform* 2015;216:50-54. [Medline: [26262008](https://pubmed.ncbi.nlm.nih.gov/26262008/)]
41. Wang H, Li J, Zhang Y, Suzuki M, Araki K. Creating personalised clinical pathways by semantic interoperability with electronic health records. *Artif Intell Med* 2013 Jun;58(2):81-89. [doi: [10.1016/j.artmed.2013.02.005](https://doi.org/10.1016/j.artmed.2013.02.005)] [Medline: [23466439](https://pubmed.ncbi.nlm.nih.gov/23466439/)]
42. Samwald M, Miñarro GJA, Boyce RD, Freimuth RR, Adlassnig K, Dumontier M. Pharmacogenomic knowledge representation, reasoning and genome-based clinical decision support based on OWL 2 DL ontologies. *BMC Med Inform Decis Mak* 2015 Feb 22;15:12 [FREE Full text] [doi: [10.1186/s12911-015-0130-1](https://doi.org/10.1186/s12911-015-0130-1)] [Medline: [25880555](https://pubmed.ncbi.nlm.nih.gov/25880555/)]
43. Samwald M, Freimuth R, Luciano J, Lin S, Powers R, Marshall M, et al. An RDF/OWL knowledge base for query answering and decision support in clinical pharmacogenetics. *Stud Health Technol Inform* 2013;192:539-542 [FREE Full text] [Medline: [23920613](https://pubmed.ncbi.nlm.nih.gov/23920613/)]
44. Tao C, Solbrig H, Chute C. CNTR0 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives. *AMIA Jt Summits Transl Sci Proc* 2011;2011:64-68 [FREE Full text] [Medline: [22211182](https://pubmed.ncbi.nlm.nih.gov/22211182/)]
45. Clark K, Sharma D, Qin R, Chute CG, Tao C. A use case study on late stent thrombosis for ontology-based temporal reasoning and analysis. *J Biomed Semantics* 2014;5(1):49 [FREE Full text] [doi: [10.1186/2041-1480-5-49](https://doi.org/10.1186/2041-1480-5-49)] [Medline: [25540680](https://pubmed.ncbi.nlm.nih.gov/25540680/)]
46. Zhu Q, Tao C, Ding Y, Chute C. Time-related patient data retrieval for the case studies from the pharmacogenomics research network. *J Med Syst* 2012 Nov;36 Suppl 1:S37-S42 [FREE Full text] [doi: [10.1007/s10916-012-9888-1](https://doi.org/10.1007/s10916-012-9888-1)] [Medline: [23076712](https://pubmed.ncbi.nlm.nih.gov/23076712/)]
47. Hu Z, Li J, Zhou T, Yu H, Suzuki M, Araki K. Ontology-based clinical pathways with semantic rules. *J Med Syst* 2012 Aug;36(4):2203-2212. [doi: [10.1007/s10916-011-9687-0](https://doi.org/10.1007/s10916-011-9687-0)] [Medline: [21445676](https://pubmed.ncbi.nlm.nih.gov/21445676/)]

48. Kim H, Park H, Min YH, Jeon E. Development of an obesity management ontology based on the nursing process for the mobile-device domain. *J Med Internet Res* 2013;15(6):e130 [FREE Full text] [doi: [10.2196/jmir.2512](https://doi.org/10.2196/jmir.2512)] [Medline: [23811542](https://pubmed.ncbi.nlm.nih.gov/23811542/)]
49. Jung H, Park H, Song T. Ontology-Based Approach to Social Data Sentiment Analysis: Detection of Adolescent Depression Signals. *J Med Internet Res* 2017 Jul 24;19(7):e259 [FREE Full text] [doi: [10.2196/jmir.7452](https://doi.org/10.2196/jmir.7452)] [Medline: [28739560](https://pubmed.ncbi.nlm.nih.gov/28739560/)]
50. Blois M. *Information and medicine: the nature of medical descriptions* Internet. Berkeley: University of California Press; 1984.

Abbreviations

CFTR: cystic fibrosis transmembrane conductance regulator

CNTRO: Clinical Narrative Temporal Relation Ontology

EHR: electronic health record

GO: Gene Ontology

OntoKBCF: Ontology-based Knowledge Base prototype on Cystic Fibrosis

OWL: Web Ontology Language

UMLS: Unified Medical Language System

URI: unique resource identifier

Edited by G Eysenbach; submitted 29.01.18; peer-reviewed by W Goossen, C Tao; comments to author 21.03.18; revised version received 25.04.18; accepted 19.07.18; published 21.12.18

Please cite as:

Jing X, Hardiker NR, Kay S, Gao Y

Identifying Principles for the Construction of an Ontology-Based Knowledge Base: A Case Study Approach

JMIR Med Inform 2018;6(4):e52

URL: <http://medinform.jmir.org/2018/4/e52/>

doi: [10.2196/medinform.9979](https://doi.org/10.2196/medinform.9979)

PMID: [30578220](https://pubmed.ncbi.nlm.nih.gov/30578220/)

©Xia Jing, Nicholas R Hardiker, Stephen Kay, Yongsheng Gao. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 21.12.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.