ONSTR: The Ontology for Newborn Screening Follow-up and Translational Research

Snezana Nikolic[1], Prabhu Shankar[3,1], Sivaram Arabandi[4], Akshaye Dhawan[5], Rajshekhar Sunderraman[6], Sham Navathe[7], Kunal Malhotra[7], Rani H. Singh[1,2,*]

[1]Deparment of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA
[2]Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA
[3]Emory Children's Center, Emory University and Children's Healthcare of Atlanta, Atlanta, GA, USA
[4]OntoPro LLC, Houston, TX, USA
[5]Department of Mathematics and Computer Science, Ursinus College, Collegeville, PA, USA
[6]Department of Computer Science, Georgia State University, Atlanta, GA, USA
[7]College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

Translational research in the field of newborn screening system requires integration of data generated during various phases of life long treatment of patients identified and diagnosed through newborn dried blood spot screening (NDBS).

In this paper, we describe the Ontology for Newborn Screening Follow-up and Translational Research (ONSTR). ONSTR is an application ontology for representing data entities, practices and knowledge in the domain of newborn screening short- and long-term follow-up of patients diagnosed with inheritable and congenital disorders. It will serve as a core of the data integration framework in the Newborn Screening Follow-up Data Integration Collaborative (NBSDC), designed to support Semantic Web tools and applications with the goal of helping clinicians involved in translational research. Here, we describe the ONSTR domain, our top-down bottom-up methodological approach to ontology modelling using phenylketonuria (PKU) as an exemplar, and some of the lessons learned. We provide an illustration of our ontological model of three important aspects of PKU: 1) the etiology, 2) the phenylalanine hydroxylase enzyme dysfunction underlying PKU and 3) the disambiguation of terms central to PKU appearing in the literature. In modelling of the mechanism of phenylalanine hydroxylase enzyme (PAH enzyme) dysfunction, we encountered limitations in using Gene Onotology (GO) process classes, in terms of their over-granularity and the lack of representations of process participants. As a solution to this problem and in order to accurately represent this process, we created a hybrid model of enzyme mediated biochemical reactions. This model of PKU and enzymatic reactions will serve as a prototype for modelling other inherited metabolic disorders (IMDs) and enzymatic processes of importance to clinical and translational research in the NBDS long-term follow-up domain.

This initial work provides the ontological foundation for automated reasoning, integration and annotation of data collected through the newborn screening system.

1 BACKGROUND

The newborn screening system (NBSS) is a comprehensive and multifaceted set of processes, protocols and medical services whose ultimate goal is to prevent the emergence of debilitating or fatal conditions via 1) the early identification of infants with inherited or congenital disorders, 2) the initiation of treatment while the disorder is still in its pre-symptomatic phase, and 3) the follow-up of patients' overall development while continuously improving the treatment(s) [1]. As recommended by SACHDNC[1], each year more than 98% of approximately 4 million newborns in the United States [2] are screened for 31 core conditions (primary conditions). In some states newborns are screened for additional 26 secondary disorders, for a total of 57 screened disorders [3].

The newborn screening system includes point-of-care procedures, such as screening for critical congenital heart diseases (CCHD) and hearing screening as well as dried blood spot screening (NDBS) usually performed in a centralized state-approved clinical laboratory. The vast majority of the SACHDNC-recommended screening disorders (46 disorders) belong to the group of inherited metabolic disorders that are detectable via NDBS. As a component of newborn screening system, newborn dried blood spot screening can be conceived as a collection of more or less temporally sequential processes: 1) newborn screening (proper): the detection of the presence or absence of a screened disorder, 2) short-term follow-up (STFU): prompt identification and location of affected infants, referral to specialists and diagnosis confirmation, 3) long-term follow-up (LTFU): treatment initiation, disorder management, treatment quality evaluation and improvement, and new knowledge generation [4, 5].

NDBS and STFU commonly generate a limited amount of data that are critical the for early detection and diagnosis of a screened disorder. In contrast, LTFU of patients with inherited metabolic disorders generates a multitude of data at different levels of granularity produced by a variety of measurement methods, assessments and evaluations carried out during regular follow-up visits. Due to the low incidence of IMDs, the collected LTFU data are often scant with respect to the numbers of patients' data pertaining to a particular inherited metabolic disorder and usually are stored in the point-of-care databases of various

* To whom correspondence should be addressed: rsingh@emory.edu

[1] SACHDNC, the U.S. Department of Health and Human Services Secretary's Advisory Committee on Heritable Disorders in Newborn and Children.

sizes and schemas. In order to be able to conduct quality translational research and generate new knowledge, however, a large amount of data is needed to support the validity of the research. At present, all existing LTFU data aggregation projects follow the traditional method of storing the data into relational databases without common data definitions and models thereby leading to the creation of data silos. The Semantic Web [6] capabilities, based on the standardized common formats that inherently support the aggregation and integration of data from different sources [7, 8] have not yet been explored in this domain. To fill in this gap, we are proposing a novel approach to LTFU data aggregation and integration. We initiated and are currently building the infrastructure for the Newborn Screening Follow-up and Translational Research Data Integration Collaborative (NBSDC) [9], envisioned as an integrated framework for sharing, collecting, aggregating and integrating de-identified patient data. NBSDC is the first project in the LTFU field that relies on applying the Semantic Web technologies and ontologies built to serve as a useful tool for the meaningful use of collected electronic patient data and new knowledge generation.

Inherited metabolic disorders arise when mutations of one or more genes that code for a particular enzyme produce a misfolded enzyme protein sub-units that form a dysfunctional enzyme complex which subsequently cannot catalyze certain biochemical reactions. Enzyme malfunction, consequently, causes the accumulation of particular metabolite(s) in body fluids (usually blood and cerebrospinal fluid (CSF)). If left untreated, IMDs can cause noticeable developmental delay within a couple of months after birth, and later severe intellectual disability and in some cases death. The management of patients with IMDs usually consists of nutrition therapy and/or a combination of medical and nutrition treatments, which in essence consists of a very restricted intake of a particular nutrient that cannot be catabolized in metabolic reactions involving dysfunctional enzyme(s), while concurrently ensuring adequate intakes of all other nutrients required to maintain physiological/metabolic homeostasis. To accomplish this task concurrent with following a patient's growth and development, long-term monitoring and periodic comprehensive assessments of the patient's nutrient intake, metabolic control of specific nutrients and toxic biomarkers are required on a regular basis.

Our initial ontology development efforts are focusing on the follow-up of patients with one of the screened IMDs: phenylketonuria (PKU). The choice of PKU was based on the facts that it was the first of all known IMDs to be screened for and that the nutrition treatment of PKU has the longest and most successful history among the screened IMDs, providing the richest pool of available scientific literature.

In this paper we present the Ontology for Newborn Screening Follow-up and Translational Research (ONSTR) [10], designed to provide formalized representation of the domain knowledge to support analysis of and reasoning over the data collected and integrated under the NBSDC. In the following sections we describe the ONSTR domain, ONSTR development methods and some of the results of our current ontology modeling efforts.

2 METHODS

2.1 The Scope of ONSTR

ONSTR is a structured formal representation of data entities, processes and agents involved in practices of newborn screening (proper), short- and long-term follow-up of patients with inheritable and congenital disorders and related translational research. The ONSTR domain lies at the crossroads of multiple scientific disciplines: medicine, nutrition, genetics, biochemistry, neurology, psychology, and public health. It includes 1) different types of measurement data obtained through various measurement methods, assessments and evaluations (including full blood count, genotyping, sequencing and imaging data, nutrient intakes, anthropomorphic measurements, psycho-cognitive evaluations, data about population samples, etc.) and 2) various kinds of processes, procedures, protocols and agents involved in the detection, diagnosis and treatment of screened disorders. In terms of the granularity level, the ONSTR domain covers representations of entities on a molecular level (such as pyridoxine and tyrosine), molecular complexes (e.g., phenylalanine hydroxylase complex and ferritin complex), cells, human anatomical structures, organ systems and patient populations (e.g., population of phenylketonuria patients).

2.2 ONSTR Development Methodology

Our ontology development method generally follows the phases outlined in Uschold and King 1995 [11]. Along with following the principles of ontological realism [12], we adopted the dual *top-down bottom-up* approach in our ontology capture phase [11]. Applying the *top-down* method, with the goal of creaing an exhaustive list of terms and relations used in our domain of interest (domain terminology), we performed an extensive search of relevant literature and electronic resources, from government documents and websites, through biochemistry and medical textbooks to single patient case studies. We also surveyed existing ontologies and controlled vocabularies mostly available at the NCBO BioPortal [13, 14] and identified the terms and relations in relevant source ontologies that are potentially importable into the ONSTR. Since data generated in everyday clinical practice and research are the data item entities to be represented in the ontology, in terms of the *bottom-up* (data-driven, practice-driven) approach, we 1) observed and analyzed the day-to-day practice of

newborn screening and patient follow-up visits and 2) included the Newborn Screening Translational Research Network (NBSTRN)-recommended data elements that are compiled by the nationally recognized domain experts and included in the Recommended Uniform Screening Panel [3].

Following the best practices of the OBO Foundry [15] and principles of ontological realism [12], ONSTR adopts the Basic Formal Ontology (BFO2) [16] as the top-level ontology and reuses classes and relations already asserted in the OBO Foundry ontologies and candidate ontologies. In most cases, ONSTR builds its hierarchies under the classes imported from 13 OBO Foundry ontologies by using the principles of Minimum Information to Reference an External Ontology Term (MIREOT) [17], and rarely directly under the BFO.

For each entity natively represented in the ONSTR, there obligatorily exists: 1) one unique identification number, 2) one term name or relation name (depending on the kind of represented entity) and 3) at most one natural language definition for given term/relation. Natural language (NL) definitions that provide the interpretation of each term and relation are encoded in the English language and have the Aristotelian *genus-differentia* form (*A is a B that C*) [12, 18]. Wherever possible, the NL definition for a given ONSTR term is drawn from one or multiple authoritative resources and/or domain-relevant literature, or in some cases, from a non-OBO Foundry ontology or controlled vocabulary, such as SNOMED CT [19].

The latest release of the ONSTR is available for browsing and downloading at the NCBO BioPortal page http://bioportal.bioontology.org/ontologies/49978.

3 RESULTS

ONSTR is a structured representation of data items, processes and agents involved in the process of newborn screening and follow-up of patients positively diagnosed with screened inheritable and congenital disorders, designed to support computational reasoning over collected data.

Even though the ONSTR domain covers entities related to newborn screening and follow-up of patients diagnosed with 56 currently screened disorders [3], most of our efforts are and will be dedicated to the modeling of practices related to the follow-up of patients with inherited metabolic disorders (IMDs) also known as inborn errors of metabolism (IEMs).

ONSTR is in the early stages of development andcurrently contains 1850 classes, 104 object properties, 4 data properties and 201 logical descriptions of levels of various complexity, with *SIQ(D)* as the current level of DL expressivity of the entire ontology [20].

3.1 PKU: From Gene to Disease

In this section we provide the illustration of our ontological representation of three interrelated aspects of PKU: 1) the etiology, 2) what is PKU in terms of recent ontological developments and 3) the enzyme dysfunction underlying PKU[2].

PKU is an inherited metabolic disease with an autosomal recessive inheritance mode [21] arising from mutations in the gene coding for the protein phenylalanine hydroxylase sub-unit (PAH monomer, PAH sub-unit). PAH sub-units are assembled into active homo-tetramer protein oligomers (PAH complexes) that catalyze the conversion of amino acid L-phenylalanine (Phe) into L-tyrosine (Tyr). This kind of Phe-to-Tyr conversion is the major and the only clinically important pathway of catabolism of phenylalanine in humans [21]. There are currently 548 known mutations in the human PAH gene [22, 23] producing a wide spectrum of PKU phenotypes [21]. Most frequently, the PAH genotype is compound heterozygous, with a relatively small number of homozygous cases. The nature and the combination of paternal and maternal mutated alleles in a child's PAH genotype have substantial functional and clinical significance [21]. If the exact combination of mutations of the inherited PAH alleles is known, it is possible to predict the patient's Phe tolerance and determine the best possible personalized treatment. To ontologically represent the etiology of PKU and to define in ontological terms what PKU is, we have used terms imported from several OBO Foundry ontologies: SO [24], GO [25], PR [26], FMA [27], CL [28], ChEBI [29] and OGMS [30] as well as classes asserted in the ONSTR. Figure 1 (following page) provides the illustration of our ontological model of the three aforementioned aspects of PKU. The top portion of the diagram focuses on the genetic basis of PKU and represents the most frequently occurring clinical scenario in which PKU has been inherited in the heterozygous compound state (i.e., two copies of mutated but different PAH alleles), under the assumption that the defective PAH monomers have an intact tetramerization domain (i.e., can form PAH complexes) and with the omission of the expression of PAH genes in kidney epithelial cells.

[2] In the text hereafter, ONSTR native terms and relations will be represented in italic typeface (e.g., *nutritional item*). Terms and relations imported from other ontologies will also be in italic typeface, but prefixed by the source ontology acronym, followed by a colon preceeding the term/relation name (e.g., *bfo:process profile*, *ogms:disorder*). The same ontology name prefixing will be used in all graphic representations, sans italic typeface.

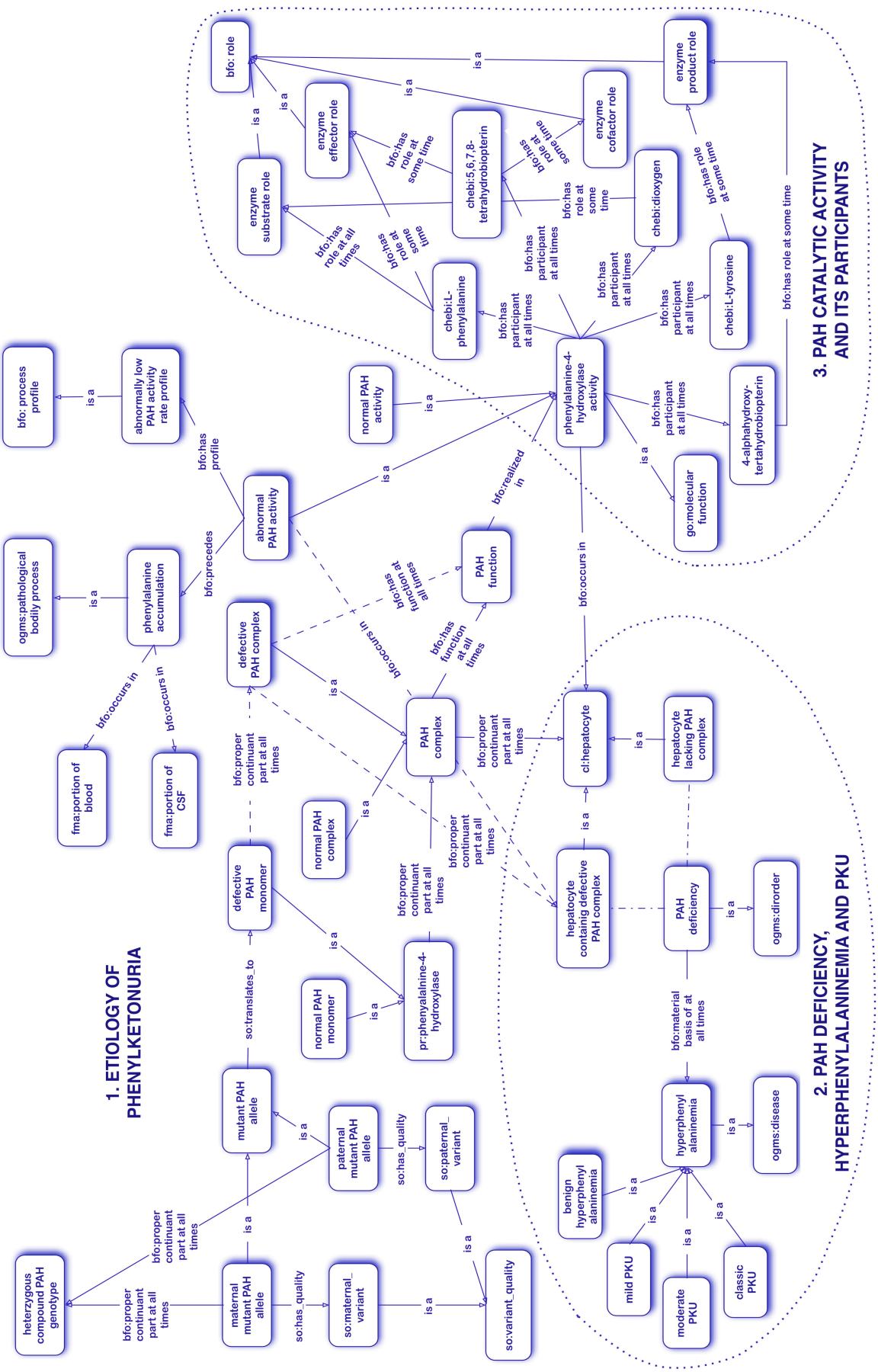

Fig. 1. Ontological representation of (1) the etiology of phenylketonuria, (2) the PAH deficiency, hyperphenylalaninemia and PKU, (3) the phenylalanine hydroxylase catalytic activity and its participants. Classes are represented by rounded rectangles. Dashed lines represent the relations holding between the instances of classes whose parents are linked through relations represented by full a line and a relation name. Dash-dot lines in the encompassed area 2 connect the classes that are identified as the *PAH deficiency* (Fig.2). For clarity, the expression of PAH genes in kidney epithelial cells and some mid-level parent classes have been omitted in the diagram. Abbreviations: PAH=phenylalanine hydoxylase, PKU=phenylketonuria, CSF=cerebrospinal fluid).

Starting from the *mutant PAH allele* and proceeding in left-to-right direction, we can read the diagram by using class and relation names along with the English verbs and connectives:

1. *mutant PAH allele so:translates_to defective PAH*
2. *defective PAH* is *bfo:proper continuant part at all times* of *defective PAH complex*
3. *defective PAH complex* is *bfo:proper continuant part at all times* of *hepatocyte containing defective PAH complex* and *bfo:has function at all times* the *PAH function*
4. *PAH function* is *bfo:realized in PAH activity* that *bfo:ocurrs in* a *cl:hepocyte*
5. *abonrmal PAH activity bfo:has profile abnormally low PAH activity rate profile* and *bfo:precedes phenylalanine accumulation* that *bfo:occurs in* the *fma:portion of blood* and in the *fma:portion of cerebrospinal fluid.*

In the literature, the above mentioned accumulation of phenylalanine in the the blood and the CSF are commonly called hyperphenylalaninemia (HPA), while PAH deficiency usually refers to the malfunction of PAH complex (PAH enzyme). These terms, however, are very polysemous across scientific literature and clinical practice. The nouns "hyperphenylalaninemia" and "phenylketonuria" are frequently used interchangeably, and often the only discernable difference between the meanings of the two could be gleaned from the clinical practice where they commonly refer to two different ranges of phenylalanine concentrations in the blood. For that reason, by relying solely on the usage of the terms, it is not easy to tell whether hyperphenylalaninemia and phenylketonuria are two separate medical conditions or one single condition with two different degrees of severity [31]. Similarly, the noun phrase "PAH deficiency" often refers to several different scenarios: 1) the total absence of PAH in the organism [21, 31], 2) the situation when PAH enzyme exists in the body, but due to its abnormally low activity, the Phe accumulation ensues [21, 31] and 3) the complete inability of the existing PAH enzyme to catalyze the conversion of phenylalanine to tyrosine [31]. In an attempt to disambiguate these terms and to answer the basic ontological question: Which is that refered by the terms hyperphenylalaninemia, PKU and PAH deficiency, we applied disciplined ontological reasoning along with following the OGMS description of diseases and disorders. In OGMS, 1) diseases are a type of *bfo:dispositions* (realizables), 2) disorders are a type of *bfo:material entity* and 3) diseases arise from disorders. By following the OGMS approach and taking into account what has been modeled so far, we could infer that *PAH deficiency* can be picked out as both the *defective PAH complex* and the *hepatocyte lacking PAH complex*. In other words, we end up with two different entities on two different levels of granularity that can both be identified as *PAH deficiency*. For level uniformity's sake and because *defective PAH complex* is, after all, *bfo:proper continuant part at all times* of *cl:hepatocyte*, we can deduce that the disorders underlying PKU and/or HPA are 1) the *hepatocyte containing defective PAH complex* and *kidney epthelial cell containing defective PAH complex* or 2) the *hepatocyte lacking PAH* and *kidney epthelial cell lacking PAH complex* (Fig. 2).

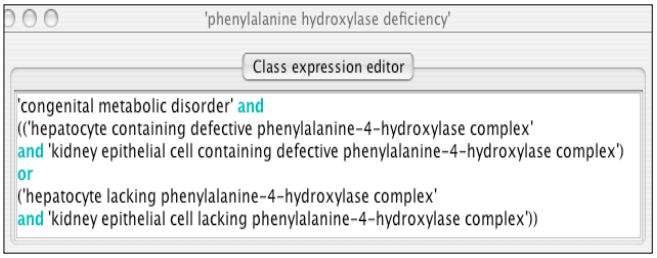

Fig. 2. Logical description for *phenylalanine hydroxylase deficiency* class, as rendered by Protégé [32] class expression editor.

Going one granularity level up, to the organ level, we may also say that *PAH deficiency* can be considered as an organ-level disorder, that is, it may be identified as a disorder of liver and kidneys (a disorderd liver or disordered kidneys), whereby liver and kidneys contain defective PAH complex, and/or liver and kidney lack PAH complex. Consequently, in accordance with OGMS, hyperphenylalaninemia and PKU may be both picked out as sub-types of *ogms:diesase*, and both as the sub-types of one single disease. The rationale for such an ontological classification is threefold: 1) Based on clinical practice both hyperphenylalaninemia and PKU arise from the *PAH deficiency* as the underlying disorder, 2) both have essentially the same disease phenotype (i.e., abnormally high Phe and abnormally low Tyr concentrations in the blood and the CSF), and 3) both are preventable from realizing/manifesting (*sensu bfo:realizable entity*) by application of the same kind of nutritional and medical treatments.

The process of metabolic conversion of Phe to Tyr by the PAH enzyme as well as the chemical entities involved in this process are of significant importance in clinical practice covered by the ONSTR domain. Tetrahydrobiopterin (BH4), for instance, is a bio-synthesized molecule that serves as a PAH enzyme cofactor and effector (inhibitor) [21], but is also used as a pharmacological chaperone in the treatment of patients whose PAH genotype makes such supplementation therapy feasible [33, 34]. To accuratelly represent and capture all relevant aspects of this process, in the true spirit of re-use, we turned to OBO Foundry ontologies, primarily to GO, PR and ChEBI. In GO, we found that there are four classes pertaining to the process of catalytic conversion of Phe to Tyr: 1) *go:phenylalanine 4-monooxygenase activity* and three other classes representing processes that stand in

part_of relationship to it, 2) *go:L-phenylalanine metabolic process*, 3) *go:tyrosine biosynthetic process, by oxidation of phenylalanine* and 4) *go:tetrahydrobiopterin catabolic process*) [35]. Furthermore, the formula stated in the NL definition of *go:phenylalanine 4-monooxygenase activity* (*go:phenylalanine 4-monooxygenase activity*: *L-phenylalanine + tetrahydrobiopterin + O2 = L-tyrosine + 4-alphahydroxy-tetrahydrobiopterin*) indicates that the team GO appears to have defined the above four processes according to the chemicals that apparently play the most important role in the process being defined; one process per important chemical participant and one term to represent that process (molecular oxygen excluded). Bearing in mind that the ONSTR domain covers the follow-up of 46 screened IMDs and assuming that each disease minimally involves one (dysfunctional) enzyme, the number of GO classes to be imported just to capture the enzyme-mediated reactions would, as a consequence, roughly (or minimally) be 190. Moreover, even though GO's process-wise granularity seems to offer sufficient ontological material to model enzyme catalytic activities, there is still a missing link needed to connect all relevant ontological elements (PR, ChEBI and ONSTR classes) into the true-to-the-fact representation of enzymatic processes.

To solve this problem, in our attempt to provide an accurate but not too cumbersome representation of enzyme-mediated reactions, we adopted a hybrid approach to model the PAH activity as a prototype for all other enzymatic processes to be modeled. In essence, we followed the situational role based approach [36, 37] in which chemicals have different roles with respect to the biochemical process they participate in. However, in contrast to the role based model [36, 37], we adhered to the BFO interpretation of functions (*bfo:function*), in the sense that enzymes' functions of catalyzing biochemical reactions do inhere in enzymes, just as the heart function of pumping blood inheres in the heart. When enzymes cannot realize to the fullest extent their catalytic function, that is when the rate of their catalytic activity is abnormally low or zero, the IMDs develop. In comparison to enzymes, chemical entities participating in enzymatic processes other than enzymes themselves are not bearers of any specific function (*sensu bfo:function*), but have different roles depending on the biochemical process they participate in. Consequently, a chemical (which is not an enzyme type) may have different roles in different enzymatic processes, or a chemical may have the same role in different biochemical reactions. For instance, the amino acid L-tyrosine, in the case of the abovementioned PAH-catalyzed conversion of Phe to Tyr, has the role of an enzyme product, while in the case of the tyrosine 3-hydoxylase mediated reaction, L-tyrosine has the role of an enzyme substrate. Similarly, BH4 has the role of an enzyme cofactor and/or enzyme effector in several different reactions catalyzed by different enzymes: tyrosine 3-hydrohylase, tryptophan 5-hydroxylase, all three forms of nitric oxide synthase and glyceryl-ether monooxygenase [38]. In representing the roles born by different molecules (that are not enzymes and) that take part in enzymatic reactions, we essentially followed the role based approach [36, 37], as well as the IUPAC classification [39]. We defined a small sub-hierarchy of roles born by chemical entities relevant for biochemical reactions to be modeled in our ontology (Fig. 3).

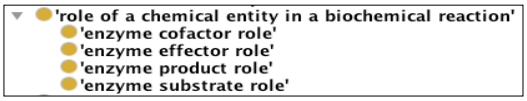

Fig. 3. Sub-hierarchy of roles chemical entities may have in biochemical reactions, as rendered by Protégé class hierarchy editor.

Additionally, since providing a logical description for a class imported from a source ontology is not a favorable practice, based on the GO class go:*phenylalanine 4-monooxygenase activity*, we resorted to defining an ONSTR class *phenylalanine-4-hydroxylase activity* for which we provided the logical description and the mapping to the source GO class. By using the above-mentioned set of roles as well as BFO2 relations, we combined classes from GO, PR, ChEBI and ONSTR to represent the *PAH activity* that unfolds in the cytosols of hepatocytes and kidney epithelial cells (excluding the cells lacking PAH complex). During the *phenylalanine-4-hydroxylase activity, PAH complex bfo:realizes* its catalytic function (*PAH function*), while the rest of the participating molecules have their roles defined in accordance with the most recent scientific literature (Fig. 1, encompassed area *2, PAH catalytic activity and its participants,* and Fig. 4, below).

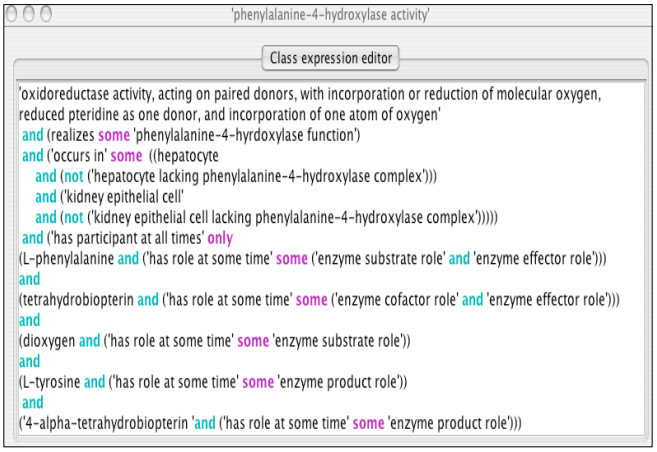

Fig. 4. Logical description for *phenylalanine-4-hydroxylase activity* class, as rendered by Protégé class expression editor.

This kind of hybrid approach to modeling of enzymatic processes proved to be very suitable for our purposes for several reasons: 1) it accuratelly represents the

functions of enzymes and roles of molecules in biochemical reactions, 2) it is fully conformant to BFO2, 3) it is flexible and re-applicable to all other enzyme-mediated processes and 4) it allows representation of interconnectedness of different and clinically important biochemical processes that are of significance for research practices pertaining to the treatment and management of IMDs.

4 CONCLUDING REMARKS

ONSTR is the first ontology of its kind primarily built to support Semantic Web tools and applications as well as reasoning over newborn screening and follow-up data. ONSTR is in the early stages of development and now covers the processes and measurements pertaining to blood metabolites relevant in newborn dried blood spot screening, growth and development follow-up and dietary treatment parameters and the genetic basis of hyperphenylalaninemia. To accurately represent the enzymatic processes, we followed a hybrid approach, a role based approach to biochemical reactions applied to imported and ONSTR native classes, but fully conformant with BFO and OBO Foundry best practices. Our future work involves testing the ontology's capabilities in integration and reasoning over real-life LTFU data collected through the newborn screening system with the goal of furthering the translational research and personalizing the care of patients with inherited metabolic diseases.

ACKNOWLEDGEMENTS

This work is partially supported by a 2009-10 HSI Seed Grant awarded by the Clinical Outcomes Research and Public Health (CORPH) Pilot Grants Program and by Georgia Tech and Children's Healthcare of Atlanta and the Southeast NBS and Genetics Collaborative Grant awarded from the Maternal and Child Health Bureau, HRSA Grant #U22MC10979.

REFERENCES

1. Lloyd-Puryear, M., Brower, A. (2010) *Long-term follow-up in newborn screening: A systems approach for improving health outcomes*. Genetics in Medicine, Vol. 12, No. 12, Dec, 2010 Supplement.
2. Rodney Howell, R., Terry, S., Olney, R., Hinton, C., Grosse, S., Eichwald, J., Cuthbert, C., Popovic, T., Gildewell. J. (2012) *CDC Grand Rounds: Newborn Screening and Improved Outcomes. Morbidity and Mortality Weekly Report.* 2012:61 (21):390-393.
3. *Recommended uniform Screening Panel of the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children.* http://www.hrsa.gov/advisorycommittee/mchbadvisory/heritabledisorders/recommendedpanel/uniformscreeningpanel.pdf
4. Hinman, A.R., Mann, M.Y., Singh, R.H. (2009) *Newborn dried bloodspot screening: mapping the clinical and public health components and activities*. Genetics in Medicine, Vol.11, No. 6, June 2009.
5. Singh, R.H., Hinman, A. R. (2010) *Newborn dried bloodspot screening: Long-term follow-up activities and information system requirements*. Genetics in Medicine, Vol. 12, No. 12, Dec, 2010 Supplement.
6. Berbers-Lee, T., Hendler, J., Lassila, O. (2001) *The Semantic Web*. Scientific American May 2001.
7. Ruttenberg, A., Clark, T., Bug, W., Samwald, M., Bodenreider, O., Chen, H., Doherty, D., Forsberg, K., Gao, Y., Kashyap, V., Kinoshita, J., Luciano, J., Marshall, S.M., Ogbuji, C., Rees, J., Stephens, S., Wong, G.T., Wu, E., Zaccagnini, D., Hongsermeier, T., Neumann, E., Herman, I. and Cheung K-H. (2007) *Advancing translational research with the Semantic Web*. BMC Bioinformatics 2007, 8(Suppl 3):S2.
8. Dumontier, M. (2010) *Building an effective Semantic Web for health care and the life sciences*. Semantic Web 1 (2010) 1-5
9. *Newborn Screening Follow-up Data Integration Collaborative.* Home page: https://nbsdc.org
10. Shankar, P., Nikolic, S., Arabandi, S., Navathe, S., Malhotra, K., Singh, R.H. (2013) *Developing the Ontology for Newborn Screening Follow-up and Translational Research.* Abstract, 2013 Joint Meeting of the Newborn Screening and Genetic Testing Symposium and the International Society for Neonatal Screening, Atlanta GA, May 2013.
11. Uschold, M., King, M. (1995) *Towards a methodology for building ontologies*. AIAI-TR-183, July 1995.
12. Smith, B., Ceusters, W. (2010) *Ontological realism: A methodology for coordinated evolution of scientific ontologies.* Applied Ontology, 2010 November 15; 5(3-4): 139-188.
13. *National Center for Biomedical Ontology (NCBO)*. Home page: http://www.bioontology.org/
14. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., Musen, M.A.(2009) *BioPortal: Ontologies and Integrated Data Resources at the Click of aMouse*. Nucleic Acids Res. 2009(37)W170-3
15. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungal, C. J., the OBI Consortium, Leontis, K., Rocca-Serra, P., Ruttenberg, A., Sansone, S-A. Scheuermann, R. H., Shah, N., Wetzel, P. L., Lewis, S. (2007) *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. Nat Biotech, 2007 November; 25(11):1251.
16. *BFO 2 Reference*. Available at http://bfo.googlecode.com/svn/release/2012-07-20-graz/BFO2-Reference-for-html.htm.
17. Courtot, M., Gibson, F., Lister, A., Malone, J., Schober, D., Brinkman, R., Ruttenberg, A. (2009) *MIREOT: the Minimum Information to Reference an External Ontology Term* . Nature Precedings, Aug 2009.
18. Berg, J. (1983) *Aristotle's Theory of Definition*. ATTI del Convegno Internazionalle di Storia de la Logica, San Gimignano, 4-8 Dec.1982.
19. *Systematized Nomenclature Of Medicine Clinical Terms (SNOMED CT)*. Home page: http://www.ihtsdo.org/snomed-ct/
20. Baader, F., Horrocks, I., USattler. (2007) Chapter 3 Description Logics. In van Harmelen F., Lifschitz, V., Porter, B. (2007) Eds Handbook of Knowledge Representation. Elsevier, 2007.
21. Blau, N., Burton, B. K., Theoeny, B., van Spronsen, F. J., Waisbren, S. (2010) *Phenylketonuria and BH4 deficiencies.* UNI-MED Verlag AG, Bremen., Germany.
22. Scriver C.R. (2007) *The PAH gene, phenylketonuria, and a paradigm shift*. Hum. Mutat.2007;28:831-845.
23. *The Phenyalanine Hydroxylase Locus Knowledgebase (PAHdb)*. Home page: http://www.pahdb.mcgill.ca/

24. Eilbeck, K., Lewis, S.E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., Ashburner, M. (2005) *The sequence ontology: a tool for the unification of genome.* Genome Biol 2005, 6:R44
25. The Gene Consortium, Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. (2000) *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000 May; 25(1): 25-29.
26. Natale, D., Arighi, C.N., Baker, W.C., Blake, J.A., Bult, C.J., Caudy, M., Drabkin, H.J., D'Eustachio, P., Evsikov, A.V., Huang, H., Nchoutmboube, J., Roberts, N.V., Smith, B., Zhang, J., Wu, C.H. (2010) *The Protein Ontology: a structured representation of protein forms and complexes.* Nucleic Acid research, 2001 Vol. 39, Database Issue D539-545.
27. Rosse, C. Mejino, J.L.V. (2003) *A Reference Ontology for Bioinformatics: The Foundational Model of Anatomy*. JBiomedInfo, 36:478-500.
28. Bard, J., Rhee, S.Y., Ashburner, M. (2005) *An ontology for cell types.* Genome Biology 2005, 6:R21.
29. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M. and Ashburner, M. (2008) *ChEBI: a database and ontology for chemical entities of biological interest*. Nucleic Acids Res. 36, D344-D350.
30. Scheuermann, R.H., Ceusters, W., Smith, B. (2009) *Toward an Ontological Treatment of Disease and Diagnosis.* Proceedings of the 2009 AMIA Summit on Translational Bioinformatics, San Francisco, CA, p.116-120.
31. *The Online Metabolic and Molecular Bases of Inherited Metabolic Diseases*. Valle, D., Volegnstein, B., Kinzler, K.W., Atonarakis, S.E., Ballabio, A., Sly, W.S., Scriver, C.R. Eds.
32. *Protégé: Ontology and knowledge base editor*. Home page: http://protege.stanford.edu/
33. Singh, R.H., Quirk, M.E. (2011) *Using change in plasma phenylalanine concentrations and ability to liberalize diet to classify responsiveness to tetrahydrobippterin in patients with phenylketonuria*. Molecular Genetics and Metabolism 104 (2011) 485-491.
34. Quirk, M.E., Dobrowolski, F. S., Nelson, B.E., Coffee, B., Singh, R.H. (2012) *Utility of phenylalanine hydoxylase genotype for tetrahydrobiopterin responsiveness classification in patients with phenylketonuria.* Molecular Genetics and Metabolism 107(2012)31-36.
35. The Gene Ontology Consortium. (2011) *The Gene Ontology: enhancements for 2011*. Nucleic Acid Research, 2011, Vol 40.
36. Dumontier, M. (2008) *Situational modeling: Defining molecular roles in biochemical pathways and reactions.* Proc. OWLED, 2008.
37. Boelling, C., Dumontier, M., Weidlich, M., Holzhuetter, H-G. (2012) *Role-based representation and inference of biomchemical processes.* ICBO 2012:3rd international Conference on Biomedical Ontology.
38. Thoeny, B., Auerbach, G., Blau, N., (2000) *Tetrahydrobiopterin biosynthesis, regeneration and functions.* Biochem. J., 2000, 347, 1-16
39. *The IUPAC Gold Book.* Available at: http://goldbook.iupac.org/PDF/goldbook.pdf