Ontology of cancer related social-ecological variables

Dharani K. Balasubramanian[1,*], Jamillah Z. Khan[1], Jiang Bian[1], Yi Guo[1], William R Hogan[1], Amanda Hicks[1]
[1]Department of Health Outcomes and Policy, University of Florida, USA

ABSTRACT

Several social-ecological (SE) factors affect human behavior. Analysis of these factors is an integral part of behavior research. An efficient method of scrutinizing these predictors is multilevel analysis. Social Ecological Model (SEM) is a multilevel framework that helps to capture all the variables at five levels: individual, interpersonal, organizational, community, and policy. This work aims to develop a reference ontology with classes that correspond to SE predictors that influence cancer diagnosis, beginning with the individual level of SEM. This ontology is built with an aim to aid data integration in order to carry out multilevel analysis of the integrated data. The broad hypothesis is that, if all the variables gathered from various sources and at different levels of the SEM are configured in an ontology, there will be enough information to identify and visualize association between these variables and health outcomes. This work is focused on 13 SE variables which were first identified by performing a scoping literature review. Manually curated terms corresponding to these variables were aligned with existing ontology classes. The ontology of cancer related social-ecological variables (OCRSEV) is built upon the Basic Formal Ontology 2.0 (BFO 2.0) and conforms to Open Biomedical Ontologies (OBO) Foundry's best practices. Future work is planned to extend the ontology for variables in other levels of SEM and map the PCORnet Common Data Model (PCORnet CDM) data and other relevant data with these variables in the ontology.

1 INTRODUCTION

One of the greatest advances in explaining the predictors of disease is the identification of social and psychological conditions of individuals (Chapter 4: Social Risk Factors, 2001). These conditions may influence the morbidity and mortality directly through physiological processes and indirectly through behavioral pathways. People living in areas with socioeconomic disadvantages and cultural barriers such as lack of health insurance, low income level, and negative opinions toward cancer screening are more likely to be diagnosed with cancer at late-stage (Wang, Luo and McLafferty, 2010). Treatment is less effective and there is a lower chance of survival with those who have late-stage diagnosed. It is critical to identify and enhance methods to detect cancer at an earlier stage. Earlier cancer diagnosis would save tens of millions of dollars annually and increase chances of effective treatment and survival (Why is early diagnosis important?, 2015).

Multilevel analysis in public health research is a statistical strategy which involves parallel examinations of individual and group level factors. This method has a hierarchical approach with multiple levels that classify factors. This approach can be used in improving health outcomes by taking steps to intervene on the variables of influence (Diez-Roux, 2000). By integrating data with the help of OCRSEV (K. Balasubramanian, 2017), multilevel analysis of SE predictors can be performed.

Social Ecological Model (SEM) is a conceptual model that has been in use in the public health community as the foundation of multilevel intervention design and implementation (Moore A *et al.*, 2015). SEM helps to explore the predictors of a disease at multiple levels and analyze them. There are several SE factors contributing to a wide range of health disparities in society (Adler & Newman, 2002). In this work, we refer to the SEM of health promotion (CDC – Social Ecological Model – CCRP, 2017) to stratify the SE variables of late-stage cancer diagnosis (LSCD). The SEM framework has five levels: individual, interpersonal, organizational, community, and policy. The theory of SEM model implies that the individual is the target of this interrelated system, placing the individual in the center of the model and all other levels representing SE variables around the individual level in a concentric manner (Fig. 1). When each of those variables is identified and analyzed, public health activities can be implemented at each level, thereby increasing synergies of intervention to improve the overall health outcomes of individuals.

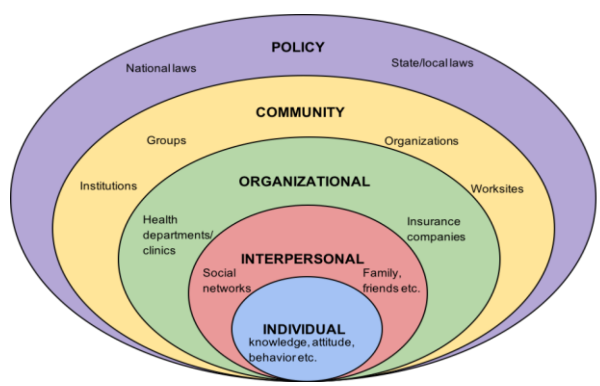

Fig. 1. Social Ecological Model, adapted by the CDC (CDC – Social Ecological Model – CCRP, 2017).

[*]To whom correspondence should be addressed: kbdhaar@ufl.edu

The structure of this ontology is based on Web Ontology Language 2 (OWL 2) models for different individual level SE variables and their relations with the SEM. For our use-case, and because this study focused on individual-level variables affecting cancer diagnosis, we investigated five databases with Florida patient data and selected one that relates the most to the individual factors of the SEM. The PCORnet Common Data Model (CDM) conveys a specification that defines a standard organization and representation of data for the PCORnet Distributed Research Network. (PCORnet Common Data Model – PCORnet, 2017). This CDM holds patient-centered data that include many individual-focused variables such as demographics and vitals. With this model, relevant data from any PCORnet database can be mapped with its corresponding classes in the ontology.

Semantic data integration (SI) combines heterogeneous data from various sources and integrates them by leveraging the semantic content that is embedded in these sources (Livingston *et al*., 2015). It brings together data from various sources by relating the true meaning of data from one source with another. Ontology mapping is key to SI of data which helps to identify similarities between ontologies and determine the concepts that represent these correspondences in order to enable reasoning. Other data integration methods may not use semantic standards, which are mandatory to define the meaning of the data and finding its roots (Noy, 2004). Our ontology can be extended for other levels of SEM and linked with corresponding data from different databases thereby providing semantic data integration.

The Ontology for Medically Related Social Entities (OMRSE) (Hogan, Garimalla, & Tariq, 2011)(Hicks *et al*., 2016) covers the domain of social entities related to healthcare and is similar to OCRSEV as it includes terms describing SE variables such as demographics, but different in the sense that it does not follow a multi-level model approach. OCRSEV is useful in knowledge representation of SE variables and specifying their level in SEM.

Scope of the Ontology: The ontology will include terms that are related to SE variables of five types of cancer diagnosis namely breast, cervical, lung, prostate and colon/colorectal. The ontology will focus on some of the key variables in the individual level of SEM. The specific aims of this project are:

- To collect SE variables related to cancer and those which act as possible determinants of late-stage cancer diagnosis based on a scoping literature review
- To develop an ontology for SE factors responsible for cancer diagnosis at the individual level of SEM.
- To include those variables that match with the individual level variables in PCORnet CDM.

2 METHODS

This work involved performing a scoping literature review (Arksey *et al*., 2007) to collect all the SE variables from different sources such as PubMed, Web of Science, Google Scholar, *etc*., that affect cancer diagnosis. A total of 79, 29, 59, 62 and 36 variables were collected for cancer types breast, lung, prostate, colon/colorectal and cervical respectively from a total of 43 articles. The variables from the literature were then extracted and stratified into the five levels of SEM.

As a preliminary step in constructing the ontology, the most common individual-level variables were selected to be used in the development of our ontology. A total of 13 variables were selected: comorbidity, age, blood pressure, body mass index range, education, employment, income status, medical cost, socioeconomic status, tobacco-use, ethnicity, race and gender. The references for these variables can be found on GitHub (K. Balasubramanian, 2017). Of these variables, gender, ethnicity, race and tobacco-use can be found in PCORnet CDM.

The competency questions (CQ) used in the development of this ontology are:

(1) What are some diseases that are cancer?
(2) What are some parts of SEM that are levels?
(3) What are some variables represented by SEM?
(4) What are some variables represented by the SEM at the individual level?
(5) What are some variables represented by the SEM at the individual level, that are also present in PCORnet CDM database?
(6) What are some variables represented by the SEM at the individual level, that are also present in PCORnet CDM database and are identity data?

Based on these competency questions, reusable classes and object properties from existing ontologies relating to our 13 variables and the five cancer types were identified using Ontobee (Ontobee, 2017) and BioPortal (NCBO BioPortal, 2017). The ontology was developed in Protégé 5.1 (Research, 2017) and conforms to OBO Foundry's (Smith *et al*., 2007) best practices. The upper-level ontology is BFO 2 (Grenon & Smith, 2004; Arp *et al*., 2015). Ontofox (Xiang *et al*., 2010) was used to import the ontologies and the individual terms for this project.

Imported Ontologies: The structure of the ontology was heavily influenced by OMRSE. OMRSE utilizes BFO as its upper-level ontology. OMRSE has many classes that are reused in OCRSEV. The Vital Sign Ontology (VSO) (Goldfain *et al*., 2011) also uses BFO as its upper-level ontology and focuses on vital sign measurement data, vital sign measuring processes, and devices used to measure human vital signs. OMRSE, along with VSO were both imported ontologies. Classes related to "blood pressure" from VSO are reused in OCRSEV. The two selected ontology imports serve as a foundation for the creation of this ontology and provides the skeleton of the future work which involving the inclusion of variables beyond the individual-level.

Imported terms: Apart from entire ontologies, some classes and relations were also imported for the construction of OCRSEV. A total of 70 terms were imported from the Human Disease Ontology (DOID) (Kibbe *et al*., 2014) for the five types of cancer (cervical cancer, colon cancer, breast cancer, lung cancer, prostate cancer). The terms were arranged in BFO 2 following the hierarchy of the Cell Line Ontology (CLO) (CLO, 2017). "Body Mass Index" was imported from Clinical Measurement Ontology (CMO) (Shimoyama *et al*., 2012; Smith *et al*., 2013). To determine where to place this class in BFO 2, the same class from Taxonomy for Rehabilitation of Knee Conditions (TRAK) (Button *et al*., 2013) was referred to.

Existing definitions for variables such as comorbidity, age, weight, and height encouraged us to reuse these definitions in place of creating our own definitions. "Comorbidity" was imported from the Ontology for Minimum Information About Biobank data Sharing (OMIABIS) (Brochhausen *et al*., 2013), and qualities such as "height" and "weight" were imported using the Phenotypic Quality Ontology (PATO) (WG, 2017). "Age" used for modeling the created term "age range variable" was imported from Ontology for Biomedical Investigations (OBI). Since OCRSEV is constructed with numerous variables and categorical variables, classes "variable" which is a subclass of "data item" and "categorical variable" which is a subclass of "variable" were reused from Apollo Structured Vocabulary (Apollo SV) (Hogan *et al*., 2016).

"Hypertension" and its related classes were imported from Obstetric and Neonatal Ontology (ONTONEO) (Farinelli *et al*., 2016) as they are useful creating class restrictions in the model for one of the created classes "blood pressure range variable". "tobacco", "tobacco material", "chewing tobacco behavior", "smoking behavior" and related classes were imported from Ontology for Biobanking (OBIB) (Brochhausen *et al*., 2016) as they were useful in developing the model for the created class "individual tobacco use variable". Since the ontology uses PCORnet CDM database variables, the class "database" was imported from Eagle-I Research Resource Ontology (ERO) (Vasilevsky *et al*., 2012).

Created Classes: SE variables were created as subclasses of categorical variable. The classes include "age range variable", "individual tobacco use variable", "individual education level variable", "individual employment status variable", "individual income status variable", "individual socioeconomic status variable", "individual medical cost variable", "body mass index range variable" and "blood pressure range variable". This covers nine of 13 variables (Fig. 2). Of the remaining four variables, three classes, namely, "racial identity datum", "ethnic identity datum" and "gender identity datum" were reused from OMRSE while "comorbidity" was reused from OMIABIS. "Cancer comorbidity" was created as a subclass of "comorbidity" and defined as a comorbidity with a primary diagnosis of cancer. This was required to relate comorbidity with cancer. "tobacco use role" and "tobacco use behavior" were created according to the model created for "individual tobacco use range variable". For the PCORnet database and variables, "PCORnet CDM database" was created as a subclass of "database" imported from ERO (Fig. 2) first and then, "PCORnet tobacco use range variable" was created as a subclass of "individual tobacco use range variable".

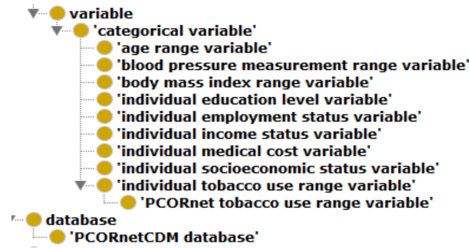

Fig. 2. Variable terms and database terms

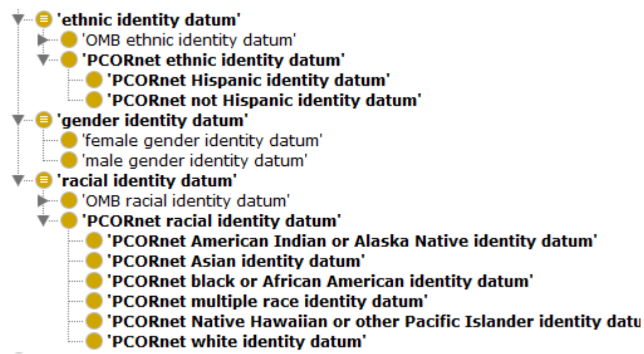

Fig. 3. Identity datum terms

The class "model" was created as a subclass of "directive information entity" in BFO following the placement of "data representational model" from Apollo SV. Since SEM of health promotion is a model that represents SE variables and is encoded in the specification of CDC's SEM, CCRP, 2015, this model is a directive information entity, similar to the "data representational model" from Apollo SV. "Multilevel model" was created as a subclass of "model" and "socio-economic model" was created as a subclass of "multi-level model". "Socio ecological model level" which is a part or a component of the socioecological model was created as a subclass of "information content entity". The five levels of socio-ecological model, namely, "socio ecological model individual level", "socio ecological model interpersonal level", "socio ecological model organizational level", "socio ecological model community level" and "socio ecological model policy level" were created as subclasses of "socio ecological model level" (Fig. 4). The class "body mass" was created as a subclass of "mass" in quality of BFO, as required by the model created for "body mass index range variable". Although "body mass" is available in Vertebrate Trait (VT) ontology (Park C *et al*., 2013), we did not reuse this class as VT does not have BFO as its upper level ontology.

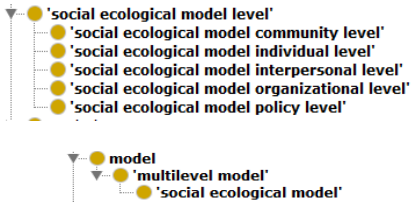

Fig. 4. Social ecological model and its levels

Classes "PCORnet ethnicity identity datum" and "PCORnet racial identity datum" and their subclasses were used from the model created for PCORowl by OneFlorida team at the University of Florida (Fig. 3) (Hicks *et al*., 2017).

Definitions: WordNet (Miller, 1990) and NCI Thesaurus (Sioutos *et al*., 2007), Wikipedia (Wikipedia, 2017), Medical Subject Headings (Medical Subject Headings, 2017), Apollo SV, PATO and other existing ontologies were referred to for crafting the definitions and modified as required based on the requirement and genus proximus (Seppälä *et al*., 2016) for all the terms created in OCRSEV. CDC's SEM for CCRP and Wikipedia were referred for crafting definitions of SEM and its levels. SEM levels' definitions are in Table 1.

Class	Genus Proximus	Definition
social ecological model level	information content entity	An information content entity which is a component of social ecological model which represents various personal or environmental variables that influence individuals and are grouped together based on some common characteristic.
social ecological model individual level	social ecological model level	A social ecological model level which represents the personal variables influencing an individual, specifically containing all the variables related to the individual's attributes, including age, gender, race, ethnicity, education, income, employment, co-morbid conditions, presence of health insurance, tobacco use and year of diagnosis.
social ecological model interperso nal level	social ecological model level	A social ecological model level which represents the social variables influencing an individual or a population, specifically containing variables related to individual's relationship with family, friends, healthcare providers, community health workers or patient navigators who can influence individual's behavior or attitude.
social ecological model organizati onal level	social ecological model level	A social ecological model level which represents the social variables influencing an individual or a population, specifically containing variables related to the availability of health care, including the number of primary care physicians and health facilities available in the individual's area of residence.
social ecological model communit y level	social ecological model level	A social ecological model level which represents the social variables influencing an individual or a population, specifically containing all the variables related to the individual's area of residence, including area-level poverty, rural residency, area-level smoking and alcohol consumption rate and area-level hospital utilization rate.
social ecological model policy level	social ecological model level	A social ecological model level which represents the social variables influencing an individual or a population, specifically containing variables related to Medicare and Medicaid at population level, including proportion of population in managed care health plan and proportion of vulnerable cancer population.

Table 1 Competency Questions, Description Logic queries, and expected results for validation

Class restrictions: To create class restrictions for the created classes, object properties such as "part of", "has part", "is about", and "inheres in" were used. Some relations such as "correlated with" were imported from Relations Ontology (RO) (Mungall, 2017) to relate the qualities or roles of the variables with cancer. From the literature review, we could say that some of the variables discussed in this paper are "correlated with" the condition "cancer" (K. Balasubramanian, D., 2017). "is determined by" and "determines" were imported from The Drug-Drug Interactions Ontology (DINTO) (Herrero-Zazo *et al*., 2015) to relate the variables with qualities or other conditions. For example, since "body mass index" is a measurement of related body weight to height of an individual, we can say that "body mass index" is determined by "height" and "weight". As an overall structure, all 13 variables are a "part of" some "social ecological model level" and every "social ecological model level" is a part of some "social ecological model". Since this work focusses only on individual level, all 13 variables are a "part of" some "social ecological model individual level". All PCORnet CDM variables are a "part of" some "PCORnet CDM database". These class descriptions were necessary to answer the six competency questions.

An example of a representation of "age range variable" is given in Fig. 5. Eight of 13 variables were modeled in the same way by relating the variable with cancer and Homo sapiens. Every variable is a data item conveying information about some other entity and can be mapped with its corresponding data (instances) from database. In Fig,5, "age range variable" is a data item which conveys information about the quality "age" which "inheres in" human being. Every SE variable is represented by a level in SEM. In our example, "age range variable" is represented by "social ecological model individual level" which is a part of "social ecological model". Hence, all SE variables which are a part of some level, all levels which are a part of SEM and the SEM are information content entities in IAO. One could draw parallels between the structure of SEM in OCRSEV and the class "document" in IAO. Just as "variable" is a part of "social ecological model level" which is a part of "social ecological model" in OCRSEV, "acknowledgements section" is a part of "document part" which is a part of "document" in IAO.

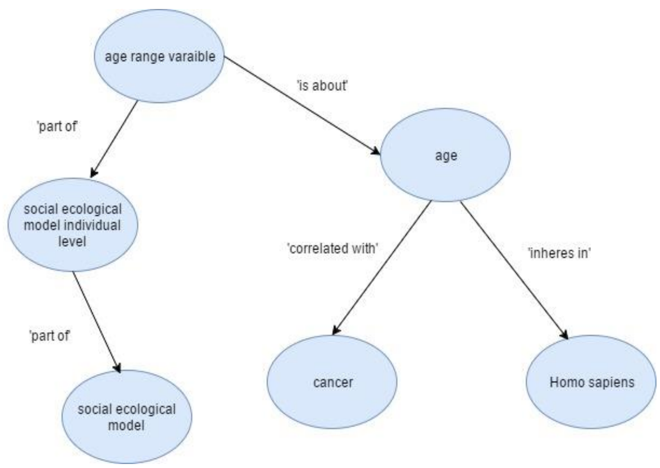

Fig. 5. RDF model for "age range variable"

The ontology was validated using Description Logic (DL) queries in Protégé 5.1. The DL queries along with the expected results are shown in Table 2.

Competency Questions	Description Logic Query	Expected Results
What are some diseases that are cancer?	disease and cancer	All 5 types of cancer (breast, colorectal, lung, prostate and cervical) and its related cancer types
What are some levels of SEM?	('part of' some 'social ecological model') and 'social ecological model level'	'socio ecological model individual level' 'socio ecological model interpersonal level' 'socio ecological model organizational level' 'socio ecological model community level' 'socio ecological model policy level'
What variables are represented by SEM?	'information content entity' and ('part of' some 'social ecological model level')	All 13 variables along with all of their subclasses
What variables are represented by SEM at the individual level?	'information content entity' and ('part of' some 'social ecological model individual level')	All 13 variables along with each of their subclasses (as this ontology is constructed for individual level of SEM)
What variables are represented by SEM at the individual level, that are also present in PCORnet CDM database?	'information content entity' and ('part of' some 'social ecological model individual level') and ('part of' some 'PCORnetCDM database')	4 variables: 'PCORnet ethnic identity datum', 'PCORnet racial identity datum', 'PCORnet tobacco use variable', 'PCORnet gender identity datum' and its subclasses
What variables are represented by SEM at the individual level, that are also present in PCORnet CDM database and are identity data?	'identity datum' and ('part of' some 'social ecological model individual level') and ('part of' some 'PCORnetCDM database'	3 variables: 'PCORnet ethnic identity datum', 'PCORnet racial identity datum', 'PCORnet gender identity datum' and its subclasses

Table 2 Competency Questions, Description Logic queries, and expected results for validation

3 RESULTS

A total of 25 classes were created of which 13 classes represent nine SE variables, five represent the five levels of SEM, three were created for SEM hierarchy and one each for PCORnet CDM specification, PCORnet CDM database, cancer comorbidity and body mass. A total of 30 class restrictions were created. The imported ontologies, classes and relations, created classes and class restrictions helped in not only answering all six competency questions but also in representing models of the 13 variables. The results were validated by running the DL queries as explained in the methods section. The actual results exactly matched with the expected results.

4 DISCUSSION

OCRSEV is an open-ended ontology that can be extended with additional variables in SEM. The principles of classification, "disjointness" and "consistent differentia" are followed for all the created classes. This project precludes the principle of "exhaustiveness" as it can include any number of variables in each level of the SEM. This project is limited to only some key SE variables of cancer and LSCD, only five types of cancer and only one multi-level model. This work is a preliminary step in providing for data integration tasks which will in turn aid in multilevel analysis of SE predictors. Future work will involve creating models for SE variables of all five levels of SEM and adding those models to the ontology. The aboutness of the variables will also be explained with regards to specifying their level in SEM as part of future work. Different databases for all the variables will be identified and respective fields will be mapped with the classes in OCRSEV. These mappings will be created using an add-on tool for Protégé 5.1 called Ontop (Calvanese *et al*., 2016).

REFERENCES

Adler, N.E. and Newman, K. (2002). Socioeconomic disparities in health: pathways and policies. Health Aff., 21(2), 60-76.

Arksey, H. and O'Malley, L. (2007). Scoping studies: towards a methodological framework. Int. J. of Social Research Methodology, 8(1), 19-32.

Arp, R., Smith, B., Spear, A.D. (2015). Building Ontologies with Basic Formal Ontology. Cambridge, MA: MIT Press.

Brochhausen, M., Fransson, M., Kanaskar, N., Eriksson, M., Merino-Martinez, R., Hall, R., Norlin, L., Kjellqvist, S., Hortlund, M., Topaloglu, U., Hogan, W.R., and Litton J. (2013). Developing a semantically rich ontology for the biobank-administration domain. J. Biomed. Semantics, 4(1), 23.

Brochhausen, M., Zheng, J., Birtwell, D., Williams, H., Masci, A.M., Ellis, H.J., and Stoeckert, C.J. (2016). OBIB-a novel ontology for biobanking. J. of Biomed Semantics, 7, 23.

Button, K., van Deursen, R.W., Soldatova, L., and Spasić, I. (2013). TRAK ontology: Defining standard care for the rehabilitation of knee conditions. J. Biomed. Inform., 46(4), 615-625.

Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M., and Xiao, G. (2016). Ontop: Answering SPARQL Queries over Relational Databases. Semantic Web Journal.

CDC - Social Ecological Model - CRCCP. (2017). Cdc.gov. Retrieved March 2017, from https://www.cdc.gov/cancer/crccp/sem.htm

Chapter 4: Social Risk Factors. Institute of Medicine. (2001). Health and Behavior: The Interplay of Biological, Behavioral, and Societal Influences. Washington, DC: The National Academies Press.

CLO. (2017). Clo-ontology.org. (2017). Retrieved April 2017, from http://www.clo-ontology.org/index.php

Diez-Roux, A.V. (2000). Multilevel Analysis in Public Health Research. Annu. Rev. Public Health, 21(1), 171-192.

Farinelli, F., Almeida, M.B., Elkin, P., and Smith, B. (2016). OntONeo: The Obstetric and Neonatal Ontology. International Conference on Biomedical Ontology, Corvallis, Oregon, USA.

Goldfain, A., Smith, B., Arabandi, S., Brochhausen, M., and Hogan, W.R. (2011). Vital Sign Ontology. Proc. of the Workshop on Bio-Ontologies, ISMB, Vienna, June 2011, 71-74.

Grenon, P. and Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation, 4(1), 69-103.

Herrero-Zazo, M., Segura-Bedmar, I., Hastings, J., and Martínez, P. (2015). DINTO: Using OWL Ontologies and SWRL Rules to Infer Drug–Drug Interactions and Their Mechanisms. J. Chem. Inf. Model., 55(8), 1698-1707.Hicks, A., Hogan, W.R., Hanna, J., Welch, D., and Brochhausen M. (2016). The Ontology of Medically Related Social Entities: Recent Developments. J. Biomed. Semantics, 7, 47.

Hicks, A., Hogan, W.R., Hanna, J. et al. (2017) PCORowl. OneFlorida Team, University of Florida.

Hogan, W. R., Garimalla, S., & Tariq, S. A. (2011). Representing the Reality Underlying Demographic Data. Paper presented at the ICBO-2011 International Conference on Biomedical Ontology: Proceedings of the 2nd International Conference on Biomedical Ontology, Buffalo, NY, USA.

Hogan, W.R., Wagner, M., Brochhausen, M., Levander, J., Brown, S.T., Millett, N., DePasse, Jay., and Hanna, J. (2016). The Apollo Structured Vocabulary: an OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation. J. Biomed. Semantics, 7(1).

K. Balasubramanian, D. (2017). kbdhaar/OCRSEV. GitHub. Retrieved June 2017 from https://github.com/kbdhaar/OCRSEV

Kibbe, W.A., Arze, C., Felix, V., Mitraka, E., Bolton, E., Fu, G., Mungall, C.J., Binder, J.X., Malone, J., Vasant, D., Parkinson, H., and Schriml, L.M. (2014). Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic Acids Res., 43(D1), D1071-D1078.

Livingston, K.M., Bada, M., Baumgartner, W.A., and Hunter, L. (2015). KaBOB: ontology-based semantic integration of biomedical databases. BMC Bioinformatics, 16, 126.

Medical Subject Headings. (2017). Nlm.nih.gov. Retrieved April 2017, from http://www.nlm.nih.gov/mesh/meshhome.html

Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. (1990). Introduction to WordNet: An On-line Lexical Database*. International Journal of Lexicography. ;3(4):235-244.

Moore A, Buchanan N, Fairley T, Lee Smith J. (2015). Public Health Action Model for Cancer Survivorship. American Journal of Preventive Medicine. 470-476

Mungall, C. (2017). Relations Ontology. Obofoundry.org. Retrieved April 2017, from http://purl.obolibrary.org/obo/ro.owl

NCBO BioPortal. (2017). Bioportal.bioontology.org. Retrieved March 2017, from https://bioportal.bioontology.org/

Noy, N. (2004). Semantic Integration: A Survey of Ontology-Based Approaches. ACM SIGMOD Record, 33(4), 65-70.

Ontobee. (2017). Ontobee.org. Retrieved March 2017, from http://www.ontobee.org/

Park C, Bello S, Smith C, Hu Z, Munzenmaier D, Nigam R et al. (2013) The Vertebrate Trait Ontology: a controlled vocabulary for the annotation of trait data across species. Journal of Biomedical Semantics.4(1):13.

PCORnet Common Data Model (CDM) - PCORnet. (2017). PCORnet. Retrieved March 2017, from http://www.pcornet.org/pcornet-common-data-model/

Research S. (2017). Protégé. Protege.stanford.edu. Retrieved April 2017, from http://protege.stanford.edu/products.php

Seppälä, S., Ruttenberg, A., and Smith, B. (2016). "The functions of definitions in ontologies," in 9th International Conference on Formal Ontology in Information Systems (FOIS 2016), Annecy, France, July 6-9 forthcoming.

Shimoyama, M., Nigam, R., McIntosh, L.S., Nagarajan, R., Rice, T., Rao, D.C., and Dwinell, M.R. (2012). Three Ontologies to Define Phenotype Measurement Data, Front. Genet., 3, 87.

Sioutos, N., de Coronado, S., Haber, M.W., Hartel, F.W., Shaiu, W-L., and Wright, L.W. (2007). NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. J Biomed Inform., 40(1), 30-43.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol., 25(11), 1251–1255.

Smith, J.R., Park, C.A., Nigam, R., Laulederkind, S.J., Hayman, G.T., Wang, S.J, Lowry, T.F., Petri, V., Pons, J.D., Tutaj, M., Liu, W., Worthey, E.A., Shimoyama, M., and Dwinell, M.R. (2013). The clinical measurement, measurement method and experimental condition ontologies: expansion, improvements and new applications. J. Biomed. Semantics, 4(1), 26.

Vasilevsky, N., Johnson, T., Corday, K., Torniai, C., Brush, M., Segerdell, E., Wilson, M., Shaffer, C., Robinson, D., and Haendel, M. (2012). Research resources: curating the new eagle-i discovery system. Database. 2012(0), bar067-bar067.

Wang, F., Luo, L., and McLafferty, S. (2010). Healthcare access, socioeconomic factors and late-stage cancer diagnosis: an exploratory spatial analysis and public policy implication. Int. J. Public Pol., 5(2-3), 237-258.

WG, O. (2017). Phenotypic quality. Obofoundry.org. Retrieved April 2017, from http://obofoundry.org/ontology/pato.html

Why is early diagnosis important? (2015). Cancer Research UK. Retrieved March 2017, from http://www.cancerresearchuk.org/about-cancer/cancer-symptoms/why-is-early-diagnosis-important

Wikipedia (2017). Wikipedia.org. Retrieved April 2017, from: https://www.wikipedia.org/

Xiang, Z., Courtot, M., Brinkman, R.R., Ruttenberg, A., and He, Y. (2010). OntoFox: web-based support for ontology reuse. BMC Research Notes, 3, 175.