Towards a teenager tailored ontology

— Supporting inference about the obesity-related health status —

Aleksandra Sojic[1]*, Walter Terkaj[1], Giorgia Contini[1], and Marco Sacco[1]

[1]Istituto Tecnologie Industriali e Automazione (ITIA) CNR, Milano, Italy

ABSTRACT

In this paper we outline the general framework of the ontology that captures the obesity-related features of teenagers. We present our particular choices regarding the ontology-structure, which should be capable of capturing (1) multiple perspectives used to describe an adolescent and (2) reasoning about the individual changes during the time. In the same line, we address several issues related to the modelling of normative concepts related to obesity and depict how the public health concern impacts classification of teenagers according to their phenotypes. In particular, we present a fragment of the ontology that supports inference about individuals and a personalised assessment of the obesity-related health conditions.

1 INTRODUCTION

Overweight and obesity are estimated to result in the deaths of about 320 000 people in western Europe every year [15]. The prevalence rates of obesity among children and adolescents motivated the public health organisations to engage in the promotion of a healthy life style [15]. Nonetheless, the understanding of the causal links between obesity and numerous socio-behavioural aspects of life-style is a complex task that involves multiple domains of knowledge and the analysis of heterogenous kinds of data. This problem can be addressed by adopting Semantic Web technologies, in particular ontologies, that are recognised as a convenient approach to deal with complex and heterogeneous information across various domains, enable knowledge generation via reasoning and support data interoperability [22, 10]. Several studies report on the use of ontology and semantic technologies to target obesity (e.g. [21, 20]). Scala et. al [20] developed an e-Knowledge platform, based on an OWL[1] ontology and SWRL [9] rules, classifying individuals according to the obesity level and certain medical conditions (Sarcopenia, Hypertension, Dyslipidemia, Diabete, Insulin resistance, Metabolic syndrome). Arash et. al [21] present the preliminary stage of the ontology designed to support a knowledge-based infrastructure, promoting healthy eating habits and lifestyles. While [20] focuses on adults and [21] on children, in this paper we present the current state of the ontology that captures formally obesity-related knowledge in a teenager tailored model. The main ontological structure is formalised in OWL, specifying certain generic classes that are applicable to any human being and represent explicitly the time-dependent changes in health condition (i.e. the issue that was not addressed by [21, 20]). At a later stage, the ontology should be able to support the information flow and interoperability between the technological tools and platforms employed to monitor the changes of health status, behaviour, and nutritional habits of teenagers. The initial step in the ontology development includes a multi-disciplinary analysis that considers a teenager as a dynamic agent who is constantly changing in the interaction with his environment. The development of the teenager-centered ontology is initiated within the European research project named PEGASO [2], whose main goal is the enhancement of self-awareness and motivation of adolescents towards a healthy lifestyle [6, 2, 3]. The project is driven by the public health concerns aiming at the decrease of the obesity-related risks to health [6, 14]. The target population of the project are the future adults whose behavioural habits at an early age can significantly impact their health status on a life-long scale [2]. The project includes several research initiatives and interventional strategies, most of which go beyond the scope of this paper, e.g. the development of serious games [17, 6] that will promote an healthy life style, the design of a life companion [3], the use of wearable gadgets equipped with sensors to monitor health status [3, 6], the design of mobile applications such as a diary used to record dietary habits, etc. [3, 6].

The multidisciplinary studies of the interactions between a teenager and his environment should provide a comprehensive model, i.e. the so-called *Virtual Individual Model (VIM)* [2] that will be used as a

*Corresponding author: aleksandra.sojic@itia.cnr.it

[1] http://www.w3.org/TR/owl2-overview/

[2] http://pegasof4f.eu/

theoretical framework of the PEGASO project. While the VIM captures obesity-related knowledge by the common representational means, e.g. natural language, tables and graphs (readable to competent human experts), the presented information is still implicit and it is not specified in a formal language. The VIM lacks a formal semantics that could disambiguate its terminological and ontological assumptions, structuring the concepts and relations in a comprehensive and machine-readable form. After several interviews with the domain-experts, the key targets of the ontology-model are specified to: (1) capture the health condition of an *Individual*; (2) detect personal obesity-related risk factors; (3) optimise the information structuring in order to provide a personalised feedback that can motivate behavioural changes towards a healthy lifestyle. In the following section we first outline several theoretical and practical aspects that will be relevant to describe how the preliminary ontology was designed and to justify our decisions about the structural segments of the ontology. Then, we present the current state of the ontology that captures the physical domain and classifies the health conditions based on assessment of the body constitution. Finally, we provide an example of reasoning over personal assessment of the obese condition by combining OWL and the Semantic Web Rule Language (SWRL) rules.

2 THE ONTOLOGY STRUCTURE

The framework of the ontology aims at integrating several fields of knowledge that are related to the problem of obesity. In this section we identify the fields that will be represented as the main ontology modules, each of which considers a particular aspect that is relevant for the problem. Regarding the methodological strategy of our approach, we keep in line with the tradition that considers *ontology* as an engineering artifact that is useful to model some aspects of the world. In other terms, we accept the position that in AI systems, "what exists" is that which can be represented [5, p. 908–909]. The aim of this section is to explain and justify our representational choices in the context of current scientific knowledge and goals of the work.

A cross-disciplinary approach

While dealing with the problem of obesity it is important to consider several factors such as physical inactivity, physiological dysfunction, unhealthy eating habits, social and psychological problems. In some cases, one of these aspects can be more decisive than the others causing overweight or obesity, in other cases, the disease is the result of a combination of many factors. In order to diagnose and modify the health status of an individual it is necessary to take into account his/her current condition through a comprehensive model capable of representing the human being as a whole. Such a model is an abstract representation that aims at integrating the cross-disciplinary knowledge about an individual, his/her characteristics and relationships in a broad context. It aims to comprehensively represent all the components that influence the *health status* of a teenager that is related to overweight and obesity. The model (i.e. VIM) focuses on three distinct levels in order to characterise the states of an individual: (1) the physical-physiological level, (2) the nutritional level, and (3) the psychological level. All these various perspectives constitute an integral approach to the understanding of an individual, providing the theoretical framework to determine the individual's health status in a dynamic manner. The theoretical model is considered as the initial ground that is used for the ontology development. The ontology model needs to specify further on a formal level the structure, concepts, and the most relevant relations that hold between them. Designing such an ontology model that integrates all the relevant perspectives is a challenging task. First, it involves collaboration of the experts with different backgrounds (philosophers, biomedical experts, clinicians, psychologists, computer scientists, and engineers), positioning the ontology in an interdisciplinary environment that is dealing with the terminological and conceptual problems in order to merge numerous (domain-specific) perspectives (see e.g. [22]). Second, the ontology design involves the decision on the ontology structure that will be crucial to its successful application. Currently, we consider the most convenient modularisation strategy [1, 4] to organise knowledge. At this point we have decided to divide the main ontology structure according to two criteria: (a) the criterion of *disciplinary perspectives* [6, 2, 3](specifying modules 1-5) and (b) the criterion of an *integrative view* (specifying modules 6-7). The resulting structure consists of the following modules:

1. *Physical* domain describing physical features of a teenager
2. *Physical Activity* and the related behaviour of a teenager.
3. *Physiological* domain
4. *Psycho-social* domain
5. *Alimentary behaviour*
6. *Trends of change* of an individual status that is formalised within (1-5). This module will be used to model the decisions about the interventional strategies and a personalised feedback to teenagers

7. *Commons* module to support the interoperability between the modules (1-6)

The ontology is currently at an early stage of development and has formalised only the segments relevant for the first module. The module distinguishes the relevant body features, linked to measurements and certain classes of health conditions that will be used to define the obesity-related status and potential risk factors. In the following subsection we present the links between the measurements characterising body constitution that are afterwords used to classify the corresponding health conditions.

Capturing normative concepts: assessment of obesity as a health condition

In general terms, a description of a teenager via some structural, functional, and behavioural characteristics is actually capturing aspects considered to be relevant to describe the teenager-specific phenotype. A phenotype is defined as a set of features of an organism that emerges as a result of interactions of his genetic material (specified as genotype) and the environment (see e.g.[11]). Herein, the genotype[3] of a teenager is not considered explicitly. We focus on the phenotypic features describing the class to which a person belongs as determined by the description of his/her physical and behavioural characteristics [11]. Thus, we consider that a person's phenotype belongs to the class *obese* based on his characteristics, description of which (despite of individual variations) fits to the description of an obese phenotype that is typical for every person of a certain *age*, *gender*, and with a specific *body mass index*. We define *typical* features of an *obese* phenotype in terms of a conventional agreement at the current stage of knowledge. The reference system that we use to characterise the physical features of an obese phenotype is provided by the World Health Organisation[14] and it includes the age and gender specific ranges of values, e.g. body mass index of a teenager (see [16]). Thus, we treat the description of body constitution as a specific characterisation of phenotype that is associated with *health condition*.

Figure 1 presents the hierarchy of the relevant health conditions, specifying the physical constitution that considers adiposity, body fat distribution, body mass, and central obesity. Each of the conditions is associated with a specific classification and linked to the reference values that characterise physical features relative to gender and age [16]. These classifications are distinct as they are using diverse criteria to describe a condition of body constitution.

Fig. 1. A fragment of the obesity-related classes used for the assessment of a personal health condition that is directly dependent on gender and age at the time of assessment.

For instance, the criterion of body mass (provided as body mass index [16]) in one of the classifications is used to distinguish people as belonging to one of the following groups: *obese, underweight, overweight* or *normal weight* [14]. According to the classification that considers fat distribution, a person may be classified either as *android* or as *gynoid*. In the following section we provide an example of the personalised assessment of the obese condition according to the measure of body mass index, relative to gender and age at the time of assessment.

A fragment of the ontology

An ontology as an artefact is not intended to cover the world in its entirety, but only chosen aspects of the world, on specific levels of abstraction, and for given purposes. Thus, we present here a 'simplified view of the world that we wish to represent'[5] while modelling obesity-related knowledge in a declarative formalism. We defined the scope of our universe of discourse that is relevant for the goals of our formal model, i.e. (1) capturing the physical characteristics of a *person* and then (2) evaluating the health condition specifically for a *teenager*. The former goal was reached by developing an OWL ontology, whereas the latter by defining SWRL rules. The class hierarchy of the OWL ontology consists of two key classes: `Person` and `HealthCondition`. These two classes are linked by a restriction involving the object property `isInHealthCondition`, so that a *person* can be associated with one or more health conditions (see Figure 2) while aiming at tracking the evolution of the health condition over time in a dynamic context. The classification of a generic health condition as belonging

[3] The class to which a person belongs as determined by the description of his/her physical material made up of DNA passed to the organism by his parents.

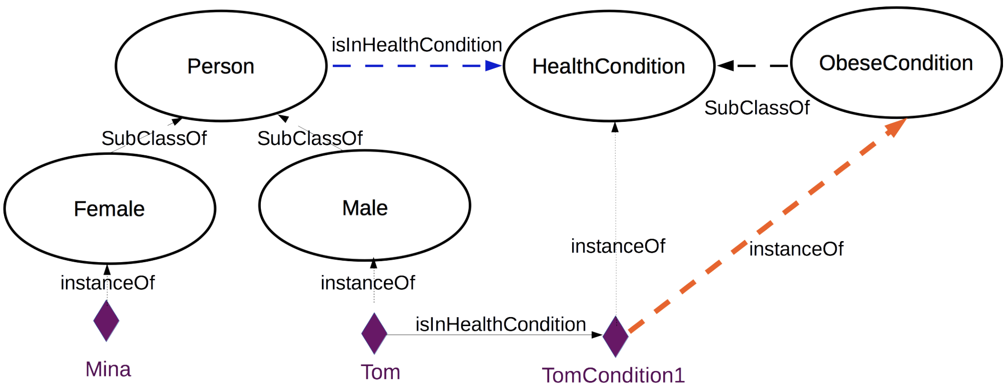

Fig. 2. Depicting the general structure of the classes and relations used to support the reasoning over instances, e.g. the orange arrow (bold dashed) stands for an inferred relation that classifies TomCondition1 as an ObeseCondition.

Rule function	Examples of the rules specified in SWRL
Calculates age of a person at the time of assessment of a health condition	HealthCondition(?h), Person(?p), isInHealthCondition(?p, ?h), isAssessedOnDateMonth(?h, ?hmonth), isAssessedOnDateYear(?h, ?hyear), isBornInMonth(?p, ?pmonth), isBornInYear(?p, ?pyear), add(?agemonthtot, ?agemonth2, ?agemonth1), divide(?ageyeartot, ?agemonthtot, 12), multiply(?agemonth2, ?ageyear, 12), subtract(?agemonth1, ?hmonth, ?pmonth), subtract(?ageyear, ?hyear, ?pyear) -> isAssessedAtAge(?h, ?ageyeartot)
Age and gender-specific assessment of a health condition	HealthCondition(?h), **Male**(?p), isInHealthCondition(?p, ?h), isAssessedAtAge(?h, ?age), isCharacterizedByBodyMassIndex(?h, ?index), greaterThanOrEqual(?age, **13.0**), greaterThanOrEqual(?index, **24.8**), lessThan(?age, **13.5**) -> ObeseCondition(?h)
Age and gender-specific assessment of a health condition	HealthCondition(?h), **Male**(?p), isInHealthCondition(?p, ?h), isAssessedAtAge(?h, ?age), isCharacterizedByBodyMassIndex(?h, ?index), greaterThanOrEqual(?age, **14.0**), greaterThanOrEqual(?index, **25.9**), lessThan(?age, **14.5**) -> ObeseCondition(?h)
Age and gender-specifc assessment of a health condition	**Female**(?p), HealthCondition(?h), isInHealthCondition(?p, ?h), isAssessedAtAge(?h, ?age), isCharacterizedByBodyMassIndex(?h, ?index), greaterThanOrEqual(?age, **13.0**), greaterThanOrEqual(?index, **26.2**), lessThan(?age, **13.5**) -> ObeseCondition(?h)

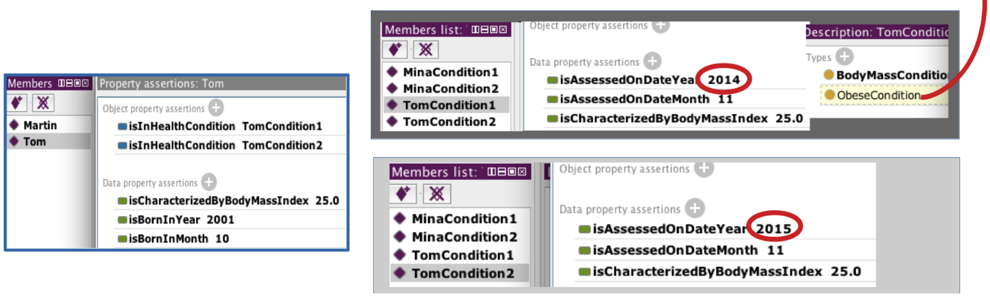

Fig. 3. The example of the rules used to support reasoning about a personalised assessment of health condition, e.g. TomCondition1 is ObeseCondition (inferred by the Pellet reasoner in Protégé.)

to one the `HealthCondition` subclasses (see Figure 1) is performed by means of SWRL rules that makes use of (1) physical (structural) and functional (metabolic, etc.) features, (2) gender, and (3) age of a person.

The *physical and functional features* are directly associated with a health condition, e.g. the body mass index is defined using the data property `isCharacterizedByBodyMassIndex` (see Figures 2 and 3).

The *gender* is defined by instantiating a `Person` as belonging to one of its subclasses, i.e. classes `Male` and `Female`.

The exact *age* of a person when the health condition is evaluated is a crucial information because the reference values for the assessment are particularly varying in the adolescence when body grows and changes [16]. In order to capture this variability that can impact on the assessment, we associate `Person` with the birth date and `HealthCondition` with the date of assessment (year and month) by using the following object properties and restrictions:

Class: `Person`
SubClassOf: (1)
isInHealthCondition **only** `HealthCondition`
isBornInYear **only** integer

isBornInMonth **only** integer
Class: `HealthCondition`
SubClassOf: (2)
isAssessedAtAge **only** decimal
isAssessedOnDateYear **only** integer
isAssessedOnDateMonth **only** integer

Having the data related to the date of birth and time of assessment, we can apply a rule modelled in SWRL[9] in order to get an age value associated with a personal condition assessment (see Figure 3), so that all the needed elaborations can be performed by a reasoning tool without needing to interface with other applications.

We tested our ontology by instantiating the above mentioned classes and then running a reasoner (i.e. Pellet reasoner plug-in for Protégé[4]) to properly classify the health conditions. The example in Figure 3 shows an instance of a boy, named Tom, who is assigned two assessments of his health condition: `TomCondition1` (assessed in November 2014), and `TomCondition2` (assessed in November 2015). Since the reference value for the assessment of an obese condition changes with age (see the reference values in Figure 3), and in our example Tom's body mass index stays unchanged, only `TomCondition1` is inferred to be *ObeseCondition*. The same reasoning can be performed with the instances of teenagers of different age and gender. We also specify the reference values relevant for assessment of other obesity-related categories of health condition (e.g. provided as measure of waist circumference) by defining a total of 74 SWRL rules. As a comparison, the work Scala et. al [20] contains approx. 40 rules.

The facts resulting from the reasoning can be saved into the ontology, thus actually enriching the knowledge. The ongoing work includes the extension of the ontology module with the formal specification of other relevant phenotypic features.

3 DISCUSSION AND FUTURE WORK

We presented the initial steps performed in the ontology design related to the PEGASO project. We depicted how a public health concern such as *obesity* impacts the decisions about the most relevant classes and relations that are used to represent phenotype of teenager and to integrate various perspectives and domain specific knowledge. Specifically, we presented the preliminary results of the ontology that (1) targets the population of teenagers, (2) formally captures physical aspects of phenotype, (3) classifies health conditions according to the physical constitution (obesity-related classes), (4) associates the condition of physical constitution with the personal assessment of the condition as age and gender dependent, (5) supports reasoning over instances, e.g. individual teenagers that may have assigned diverse conditions at different time points (age dependent assessment).

We are currently working on the design of an interlinked modular structure. Since our target is the phenotype of teenagers, in the future work we will consider to link our ontology with relevant phenotype ontologies[5]. In addition, we aim at joining the efforts in creating alignments [19, 7, 13] between our ontology and the reference terminologies and ontologies [6], as well as foundational ontologies[7].

ACKNOWLEDGEMENTS

This work has been partially funded by the EU 7th Framework Programme under the grant agreements No: 610727, "Personalised Guidance Services for Optimising lifestyle in teenagers through awareness, motivation and engagement" (PEGASO). The project is compliant with the European and National legislations concerning the user safety and privacy. We would like to thank all the partners in the project, in particular Claudio Lafortuna, Giovanna Rizzo, and Sarah Tabozzi. We would especially like to thank Laura Cruz for her contribution during the early stages of the project.

REFERENCES

[1] Camila Bezerra, Frederico Freitas, Jérôme Euzenat, Antoine Zimmermann, et al. Modonto: A tool for modularizing ontologies. In *Proc. 3rd workshop on ontologies and their applications (Wonto)*, 2008.

[2] Maurizio Caon, Stefano Carrino, Renata Guarnieri, Giuseppe Andreoni, Claudio L Lafortuna, O Abou Khaled, and Elena Mugellini. A persuasive system for obesity prevention in teenagers: a concept. In *Proceedings of the Second*

[4] http://clarkparsia.com/pellet/protege/

[5] http://bioportal.bioontology.org/ontologies/MP
http://www.human-phenotype-ontology.org/

[6] http://ncit.nci.nih.gov/,
http://www.ihtsdo.org/snomed-ct/,
http://www.nlm.nih.gov/research/umls/,
http://bioportal.bioontology.org/ontologies/HL7,
http://www.who.int/classifications/icd/en/,
http://bioportal.bioontology.org/ontologies/UO,
http://obi-ontology.org/

[7] BFO http://www.ifomis.org/bfo/, GFO [8], http://www.onto-med.de/ontologies/gfo/, DOLCE [12], GALEN [18], the Foundational Model of Anatomy (FMA), http://sig.biostr.washington.edu/projects/fm/AboutFM.html

International Workshop on Behavior Change Support Systems (BCSS2014), Padova, Italy, 2014.

[3]Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Giuseppe Andreoni, and Elena Mugellini. Pegaso: Towards a life companion. In *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management*, pages 325–331. Springer, 2014.

[4]Mathieu d'Aquin, Anne Schlicht, Heiner Stuckenschmidt, and Marta Sabou. Ontology modularization for knowledge selection: Experiments and evaluations. In *Database and Expert Systems Applications*, pages 874–883. Springer, 2007.

[5]T. R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, 43(4-5):907–928, 1995.

[6]Renata Guarneri and Giuseppe Andreoni. Active prevention by motivating and engaging teenagers in adopting healthier lifestyles. In *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management*, pages 351–360. Springer, 2014.

[7]M. Hartung, A. Groß, T. Kirsten, and E. Rahm. Effective Mapping Composition for Biomedical Ontologies. In *Proc. of Semantic Interoperability in Medical Informatics (SIMI-12), Workshop at ESWC-12*, 2012.

[8]Heinrich Herre. General formal ontology (gfo) : A foundational ontology for conceptual modelling. In Roberto Poli and Leo Obrst, editors, *Theory and Applications of Ontology*, volume 2. Springer, Berlin, 2010.

[9]Ian Horrocks, Peter F Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member submission*, 21:79, 2004.

[10]Botond Kádár, Walter Terkaj, and Marco Sacco. Semantic virtual factory supporting interoperable modelling and evaluation of production systems. *CIRP Annals-Manufacturing Technology*, 62(1):443–446, 2013.

[11]Richard Lewontin. The genotype/phenotype distinction. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2011 edition, 2011.

[12]C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. WonderWeb Deliverable D18: Ontology Library. Technical report, ISTC-CNR, 2003.

[13]Krystyna Milian, Zharko Aleksovski, Richard Vdovjak, Annette ten Teije, and Frank van Harmelen. Identifying disease-centric subdomains in very large medical ontologies: A case-study on breast cancer concepts in snomed ct. or: Finding 2500 out of 300.000. In David Riaño, Annette ten Teije, Silvia Miksch, and Mor Peleg, editors, *Knowledge Representation for Health-Care. Data, Processes and Guidelines*, volume 5943 of *Lecture Notes in Computer Science*, pages 50–63. Springer Berlin / Heidelberg, 2010.

[14]World Health Organization. *Obesity: Preventing and Managing the Global Epidemic*. IIS microfiche library. World Health Organization, 2000.

[15]World Health Organization. *European Food and Nutrition Action Plan 2015?2020*. WHO Regional Office for Europe, 2014.

[16]World Health Organization et al. Who child growth standards: methods and development: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. *Geneva: WHO*, 2006.

[17]Lucia Pannese, Dalia Morosini, Petros Lameras, Sylvester Arnab, Ian Dunwell, and Till Becker. Pegaso: A serious game to prevent obesity. In *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management*, pages 427–435. Springer, 2014.

[18]A.L. Rector and W.A. Nowlan. The galen project. *Computer Methods and Programs in Biomedicine*, 45:75–78, 1993.

[19]Alan Rector. Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. In John Gennari, Bruce Porter, and Yolanda Gil, editors, *Proceedings of the Second International Conference on Knowledge Capture (K-CAP'03), Sanibel Island, Florida, USA, Oct 23–25*, pages 121–128, New York, 2003. ACM Press.

[20]Paolo L Scala, Davide Di Pasquale, Daniele Tresoldi, Claudio L Lafortuna, Giovanna Rizzo, and Marco Padula. Ontology-supported clinical profiling for the evaluation of obesity and related comorbidities. *Studies in health technology and informatics*, 180:1025, 2012.

[21]Arash Shaban-Nejad, David L Buckeridge, and Laurette Dubé. Cope: childhood obesity prevention [knowledge] enterprise. In *Artificial Intelligence in Medicine*, pages 225–229. Springer, 2011.

[22]Aleksandra Sojic and Oliver Kutz. Open biomedical pluralism: formalising knowledge about breast cancer phenotypes. *Journal of biomedical semantics*, 3(2):1–31, 2012.