Semantic Characterization of Hypertrophic Cardiomyopathy Disease

Catia M Machado*, Francisco Couto*, Alexandra R Fernandes†, Susana Santos†, Nuno Cardim‡ and Ana T Freitas§
*LaSIGE, Departamento de Informática
Universidade de Lisboa, Lisboa, Portugal
Email: cmachado@xldb.di.fc.ul.pt, fcouto@di.fc.ul.pt
†Universidade Lusófona de Humanidades e Tecnologias, Lisboa, Portugal
Email: alexandrancrfernandes@gmail.com, sirsantoss@gmail.com
‡Hospital da Luz, Lisboa, Portugal
§Instituto de Engenharia de Sistemas e Computadores
Instituto Superior Técnico, Lisboa, Portugal
Email: atf@inesc-id.pt

***Abstract*—The application of a translational medicine approach to the study of diseases enables personalized clinical diagnosis and prognosis.**

Hypertrophic cardiomyopathy (HCM) is a disease that can benefit from such an approach, since it combines a variable clinical presentation with a genetic heterogeneity denoted by 640 known mutations, in more than 20 genes. This is a relatively common genetic myocardial disorder and the most frequent cause of sudden cardiac death in young people and athletes.

This article presents a novel semantic model representing the integration of phenotype and genotype data, mandatory for the characterization of HCM.

The model, developed in OWL Lite, comprises three connected modules: *HCM Clinical Evaluation*, *Genotype Analysis* and *Medical Classifications*. The RDF/XML representation of each module is available at https://sites.google.com/site/hcmsemanticmodel/home-1.

The lexicon of the model was based on controlled vocabularies, namely SNOMED CT, NCI Thesaurus and OCRe, with a total of 78% linked concepts.

The model will provide the basic framework for a biomedical system that will improve the diagnosis and prognosis of HCM. This improvement will be accomplished through the utilization of data mining techniques that will identify associations between the presence of certain mutations and the resulting physical traits.

***Keywords*-Translational Medicine; Semantic Web; Domain knowledge representation; Semantic Modeling; Hypertrophic Cardiomyopathy**

I. Introduction

Translational medicine can be understood as the bridging between basic and applied research, dedicated to the study of diseases. Objectively, the understanding of human diseases can result both from the integration of data obtained at the molecular and cellular level into the clinical practice (from the bench to the clinics), and from the identification of new basic research targets based on clinical observations of the disease. In a patient-oriented approach, genotype information is used in the clinical diagnosis process and translated into a personalized treatment.

The European project Advancing Clinico Genomics Trials on Cancer (ACGT) [1], [2] and the United States initiative cancer Biomedical Informatics Grid (caBIG) [3] are examples of the application of a translational medicine approach. In both cases the main objective is to provide researchers with a grid framework to share and reuse data and open source tools, aiming at the development of a post genomic research in clinical trials.

In order to implement a translational medicine approach, the genotype and phenotype data obtained from each individual need to be integrated. This data integration procedure is a complex task, since genotype and phenotype are transversal domains, and their data is normally stored under heterogeneous formats and on different locations.

Approaches based on the Semantic Web architecture have been identified as suitable for the referred type of integration task [4], [5]. The Semantic Web technologies, and in particular the standards set by the World Wide Web Consortium (W3C), such as RDF and OWL [6], enable data integration, sharing and reuse in an application- and domain-independent manner. The utilization of controlled vocabularies also contributes to the achievement of these goals, and several of such vocabularies have been developed in the biomedical domain. Examples include the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) [7], the National Cancer Institute Thesaurus (NCIt) [8], and the Ontology of Clinical Research (OCRe) [9]. SNOMED CT and NCIt are both reference terminologies describing clinical healthcare, and while SNOMED CT aims at providing a standard for the record of health care encounters, NCIt also covers vocabulary for translational and basic research, in a cancer-related setting. OCRe is a formal ontology describing human studies that contains terms related to the roles of the participant individuals, namely the term *Clinician*, which represents the role of a clinically trained medical practitioner.

The referred vocabularies can be used in the definition of domain-specific models, such as the one proposed in

this article. This model was developed to represent the translational approach to the characterization of HCM.

Examples of related work are presented in [5] and [10]. In both works Semantic Web technologies are used and cardiomyopathy is considered as a use case, but the objectives differ between them and with respect to the work presented in this article: the former work presents the integration of data for the development of a clinical diagnosis and therapeutic intervention decision support system, while the latter is related with the identification and prioritization of disease candidate genes.

HCM is an autosomal dominant genetic disease that may afflict as many as 1 in 500 individuals, and is the most frequent cause of sudden cardiac death among apparently healthy young people and athletes [11], [12]. It can manifest itself either in a sporadic form or in a familial form, and in the ultimate case the first-degree relatives of the patient may also be at risk.

Since the disease is characterized by a variable clinical presentation and onset, its clinical diagnosis is difficult prior to the development of severe or even fatal symptoms [11], [12]. Therefore, its early diagnosis is extremely important.

Currently, 640 mutations in more than 20 different genes are associated with the disease [13], [14], and the existence of a single mutation is sufficient for a positive diagnosis. However, the severity of HCM may not be the same for two individuals, even if direct relatives, since the presence of a given mutation can have a benign pattern in one individual and result in sudden cardiac death in another [11], [12].

Consequently, although genetic testing of HCM patients through dideoxy sequencing is considered a potentially valuable tool for diagnosis, it is hampered by the multiplicity of genes and mutations involved and the lack of genotype-phenotype correlations [15], [16].

In this article we propose a new semantic model covering the domain of HCM characterization through the integration of genotype and phenotype data. Based on this model a clinical decision support system will be built, providing a new framework for the improvement of the diagnosis and prognosis of this disease. With this system it will be possible to identify correlations underlying the integrated data, and possibly achieve one of the main goals of this work that is the identification of which mutations are associated with each clinical manifestation of the disease.

II. Methods

The development process of the semantic model presented herein followed the guidelines for ontology development presented by [17]. However, the execution order of the initial knowledge-engineering steps was not always the one indicated: the first step executed was, in effect, the definition of the domain and scope of the HCM model, but the second, rather than considering the reutilization of existent ontologies, was the enumeration of relevant terms for the model. This was due to the need to improve our insight of the domain to model, and resulted in the identification of the activities involved in the characterization of patients in terms of HCM, and of the data elements considered both by medical doctors and molecular biologists. These activities and data elements are represented in the Activity Map shown in Fig. 1, and were identified in collaboration with domain experts.

The next step was the representation of each category of data elements depicted in Fig. 1 (the box-like shapes) as classes in the semantic model. The only exceptions were *Muscle tissue sample* and *Transcript variant*, which will only be considered further along the development of the clinical characterization system.

To further develop the model, it was decided to focus first on the clinical evaluation elements, which include the concepts *Clinical History* (e.g. present and past symptoms) and *Exams Results* (e.g. electrocardiogram, echocardiogram). Upon collaboration with the medical experts, two more main concepts were identified, *General Characterization* of the patient (e.g. height and weight) and *Treatments* (e.g. prescription of drugs), as well as the specific data elements we could expect for each main category, that is, all possible clinical history elements, all possible exams, and so on.

Clinical History, *Exams Results*, *General Characterization* and *Treatments* were represented as siblings, but it was necessary to decide how the elements of each of those categories should be represented. The possibilities were: as properties (elements that describe features and attributes of a concept), as instances, or as subclasses of the respective main concepts. Neither of the first two options was the best solution since in that format it was not easy to maintain semantic coherence and directly relate each individual data element with, for instance, a data value or a collection date; and the last option implied that we would have instances for those subclasses, which meant that it was necessary to define the lowest level of granularity to be considered in the model. The option chosen was the representation as subclasses, and considering that the patients are the central element of the HCM model, all instances of all classes in the model represent a measurement or statement concerning a patient.

At this point we realized that the classes relative to the molecular biology analysis should have a representation independent of that of the clinical analysis, since their instances do not represent data elements concerning a patient, but rather concerning a laboratory procedure performed in a sample from a patient. This resulted in the division of the model in two modules: *Clinical Evaluation* and *Genotype Analysis*.

The development of the referred modules included the definition of the properties that describe the data, and the introduction of the restrictions necessary to define their utilization in each class definition.

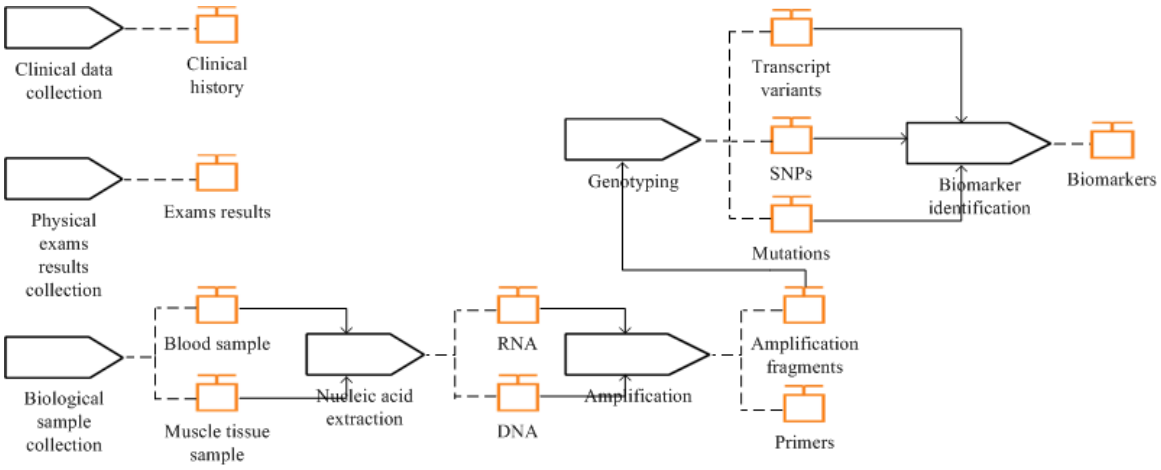

Figure 1. Schematic representation of the activities (semi-rectangular shapes) that compose the HCM characterization workflow and the correspondent generated data elements (box-like shapes). Dashed lines connect activities with generated data elements; full lines connect data elements with the activities in which they are used. The symbols used in this representation are based on the work developed by [18].

Once the backbone of the model was defined, the search for controlled vocabularies that could be reused was performed. Several were identified, and from those three were selected that contain terms necessary for the characterization of HCM patients: SNOMED CT (version 2010_01_31), NCIt (version 10.03) and OCRe (version 0.95). Since the level of complexity of these vocabularies is higher than what is needed in the HCM model, it was decided not to reuse them, but rather adopt the following approach: 1) we followed the hierarchical organization of the regions of the vocabularies containing the terms of interest; 2) we renamed the concepts in the HCM semantic model according to their equivalent terms in the controlled vocabularies; 3) we linked the HCM model concepts with terms from the vocabularies, through the URI of the latter. This approach was executed manually, and so far only for the module *Clinical Evaluation*.

SNOMED CT and NCIt were very useful for the organization of the clinical data elements. As referred above, we considered four main concepts related with the clinical characterization of the patients, represented in the model as sibling classes: *Clinical History*, *Exams Results*, *General Characterization* and *Treatments*. Upon analysis of SNOMED CT and NCIt, we performed the following modifications: the *General Characterization* elements were considered as a subtype of the *Clinical History* concept; the *Treatments* and *Exams Results* elements were considered as siblings under a parent concept *Procedure*. These are just an example of the modifications performed. In general we considered the structural organization of the term *Clinical history and observation findings* from SNOMED CT in our concept *Clinical History*, and the organization of the term *Intervention or Procedure* from NCIt in the parent concept that includes our concepts *Treatments* and *Exams Results* (named *Procedure*). There is no overlapping of terms between SNOMED CT and NCIt in the HCM model.

All concepts related to the clinical evaluation were renamed according to the selected vocabularies: *Clinical History* names as in *Clinical history and observation findings* (from SNOMED CT) and superclass *Procedure* names as in *Intervention or Procedure* (from NCIt).

A total of 78% of the concepts in the module *HCM Clinical Evaluation* were linked to the referred vocabularies, of which about 44% originate from SNOMED CT, 27% from NCIt, and 6% from OCRe. Regarding the latter, the linked terms originate from its superclass *clinical:Role*, specifically *Subject*, *Clinician* and *Health Care Site*. In respect to the module *Genotype Analysis*, it contains only one link to an external vocabulary occurring between the concept *Biological Sample* and the term *clinical:Sample*, a subtype of *clinical:Role* from OCRe. The linkage of the remaining concepts, if possible, will be performed in subsequent iterations of the model. New terms were, and will be, considered only when none of the vocabularies contain a representation of the concept to be inserted.

A third module was created to include auxiliary medical information: *Medical Classifications*. It contains medical standards used to characterize clinical elements. None of its concepts is linked to controlled vocabularies, but rather to sites where their description can be found.

The modules *Genotype Analysis* and *Medical Classifications* are connected with the module *HCM Clinical Evaluation*, which imports them.

The development of the HCM model followed a combination development process [17], in the sense that we used both a top-down and a bottom-up approach: first a top-down, when defining with the domain experts the concepts to consider, and afterwards a bottom-up, when identifying

generalizations for some of the concepts (as exemplified with the concepts *Treatments* and *Exams Results*).

The controlled vocabularies were visualized in the National Center for Biomedical Ontology BioPortal [19], [20].

The HCM domain model was developed in OWL Lite, using the Protégé-OWL editor (version 3.4.2) [21].

III. Results

Our semantic model contains the concepts, and relations between them, that represent the data needed to characterize an individual in terms of the HCM disease. In more concrete terms, the data elements to be integrated correspond to the presence of mutations in the patients' genome (genotype data) and to the clinical elements upon which the clinicians rely to provide a diagnosis (phenotype data). The latter normally include the results from physical examinations, as well as the clinical history of the individual.

The HCM model, as stated before, comprises three modules. The main module, *HCM Clinical Evaluation*, comprehends concepts associated with administrative data and with the clinical data elements necessary for the diagnosis (phenotype data). The module *Genotype Analysis* contains concepts associated with the genetic testing of biological samples (genotype data), and *Medical Classifications* is an auxiliary module containing medical standards to characterize clinical elements such as patient symptoms. The module *HCM Clinical Evaluation* imports both modules *Genotype Analysis* and *Medical Classifications* (Fig. 2).

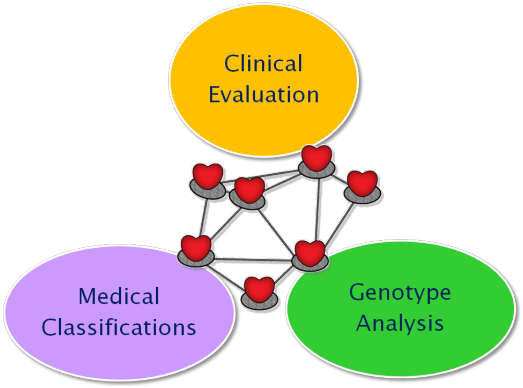

Figure 2. Schematic representation of the three modules that compose the HCM semantic model, connected by the existence of links between the data elements modeled by each module.

The RDF/XML representation of these modules is available at the following location: https://sites.google.com/site/hcmsemanticmodel/home-1. Each module can be browsed using an OWL editor such as Protégé [21]. If using this editor, the modules *Genotype Analysis* and *Medical Classifications* can be imported into *HCM Clinical Evaluation* in the Ontology Browser area of the Metadata tab.

The following subsections contain a detailed description of each module and of how they are connected.

A. HCM Clinical Evaluation

The main module of the HCM model comprises six high-level concepts (or classes), two of which pertain to administrative elements, *Health Care Site* and *Clinician*, while the remainder four refer to the subjects themselves, *Subject*, and the clinical data necessary for the diagnosis: *Clinical History*, *Procedure* and *Heart Disease*.

In total, this module contains 55 concepts and 61 properties. OWL properties can be relations between instances of two classes, or between instances of a class and a data value. For example, *hasClinicalHistory* is an example of the first type (connecting the concept *Subject* with *Clinical History*), and *hasAssociatedDate* is an example of the second (connecting, for example, a *Procedure* to its occurrence date).

Fig. 3 provides a visual representation of the high-level classes and their direct subclasses. The non-hierarchical relations between classes are not represented.

The class *Subject* corresponds to a central concept in this model, and is related to all the other concepts. It includes three subclasses: *Patient*(s) - individuals diagnosed with HCM; *Family Member*(s) - direct relatives of *Patient*(s); *Control*(s) - individuals that do not suffer from HCM.

The classes *Health Care Site* and *Clinician* do not have subclasses. *Health Care Site* refers to the institutions where the subjects receive health care services and *Clinician* to the medical doctors involved in the assessment or administration of treatment to a *Subject*.

The class *Clinical History* has five subclasses that refer to clinical elements collected upon questioning or direct examination of the subject, namely: *Cardiovascular Measurement*, *Cardiovascular Finding*, *Body Measurement*, *Resuscitated Sudden Death* and *Death*. The subclass *Cardiovascular Measurement* contains the elements *Blood Pressure* and *Pulse Rate*. *Cardiovascular Finding* contains six elements: *Angina*, *Congestive Heart Failure*, *Cardiac Auscultation Finding*, *Palpitations* and *Syncope*. *Body Measurement* includes *Weight* and *Height*. While *Resuscitated Sudden Death* has no subclasses, *Death* has two: *Sudden Death* and *Non Sudden Death*.

Regarding the class *Procedure*, its subclasses pertain to the different types of procedures to which the subject can be subjected to, namely: *Diagnostic*, *Laboratory* and *Therapeutic Procedure*(s). *Diagnostic Procedure*(s) include *Cardiac Magnetic Resonance Imaging*, *Echocardiography* and *Electrocardiographic Monitoring*. *Laboratory Procedure*(s) consist of tests carried out in biological samples, such as blood, and those considered are the *Biomarker Analysis* (in particular *Genetic Marker Analysis*) and the *Hematology Test*(s). *Therapeutic Procedure*(s) comprise the subcategories *Prescription Of Drug* and *Cardiac Procedure*, the latter including *Medical Device Implantation*, *Septal Ablation* and *Septal Myectomy*.

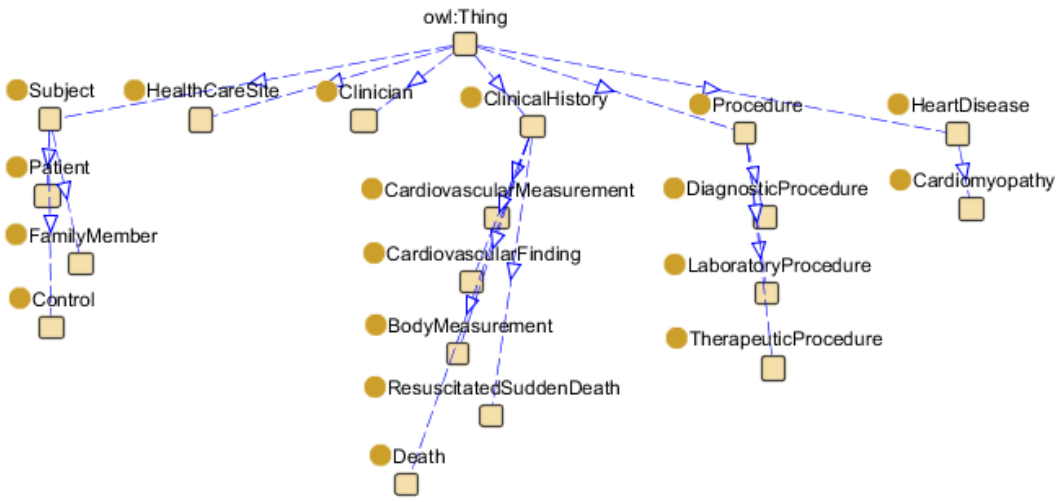

Figure 3. Graphical representation of the module *HCM Clinical Evaluation*. The figure shows the top classes of the module and their direct subclasses (when existent). The top classes are: *Subject*, *Health Care Site*, *Clinician*, *Clinical History*, *Procedure* and *Heart Disease*. The figure was obtained with Jambalaya plug-in of Protégé Editor.

The class *Heart Disease* contains cardiomyopathies (i.e., diseases of the heart's muscle), in particular *Hypertrophic Cardiomyopathy* and *Dilated Cardiomyopathy*.

For the classes *Clinical History*, *Procedure* and *Heart Disease*, the instances are records of that clinical element pertaining to a *Subject*. Considering, for example, the classes *Pulse Rate* and *Dilated Cardiomyopathy*, the instances are, respectively, a pulse rate measurement for a given *Subject* and the *Subject* to which the disease was diagnosed.

B. Genotype Analysis

The design of the *Genotype Analysis* module was oriented to the maintenance of data related to biological specimens and laboratorial activities, rather than of *Subject*'s records. It contains six high-level classes, as can be seen in Fig. 4, and a total of 39 properties. All classes are related to the process of identifying genetic markers associated with HCM in biological samples.

The activities underlying a genotype analysis involve the manipulation of *Biological Sample*(s), from which *Nucleic Acid*(s) are extracted. From the latter it is possible to obtain *Amplification Fragment*(s), which correspond to the segments of the genome to be screened for HCM-related *Mutation*(s). Each of these *Mutation*(s) is located in a specific *Gene*. All these non-hierarchical relations are represented in the model under the form of restrictions applied to the classes.

Regarding the class *Amplification Primer*, it pertains to auxiliary laboratory elements necessary for the amplification of nucleic acid fragments.

C. Medical Classifications

The module *Medical Classifications* is intended for the maintenance of data necessary for the characterization of clinical elements. Such data can be either standards or guidelines, developed to provide some degree of uniformity in the description of medical observations made by medical practitioners.

Currently, the module contains two high-level classes: *Angina Classification* and *Heart Failure Classification*, and two properties. These classes refer to functional classification systems created to assess the degree of severity of two *Cardiovascular Finding*(s), respectively angina and heart failure. Both classes have one subclass: *Angina Classification* has the classification system for angina created by the *Canadian Cardiovascular Society* [22], and *Heart Failure Classification* has the classification system for heart failure created by the *New York Heart Association* [23].

Each classification system relates the onset of the symptoms to everyday activities of the patients. In the case of angina, the Canadian Cardiovascular Society defined 5 degrees of severity, from Class 0 to Class 4. According to this classification system, a *Subject* with angina *CSS_Class1* feels chest pain associated only with strenuous exercise, while other with angina *CSS_Class4* feels chest pain at any level of physical exertion, even at rest. *CSS_Class1* and *CSS_Class4* are instances of the class *Canadian Cardiovascular Society*.

D. Bridging between modules

The module *HCM Clinical Evaluation* establishes connections with both modules *Genotype Analysis* and *Medical Classifications* (Fig. 5).

The connection with *Genotype Analysis* is made through the following non-hierarchical relations: *Subject hasBiologicalSample Biological Sample* and *Laboratory Procedure performedInBiologicalSample Biological Sample*. All elements in these relations belong to the module *HCM Clinical*

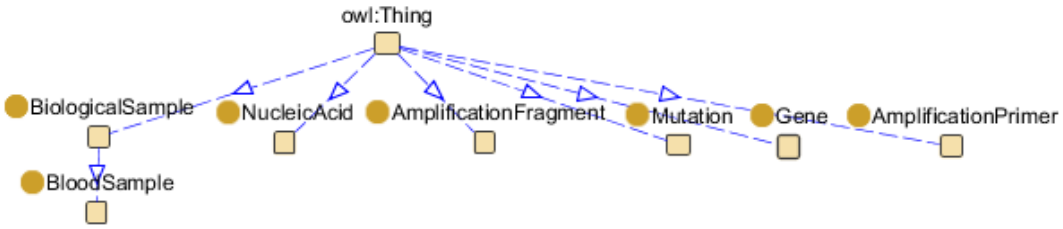

Figure 4. Graphical representation of the module *Genotype Analysis*. The figure shows the top classes of the module, namely: *Biological Sample* (with subclass *Blood Sample*), *Nucleic Acid*, *Amplification Fragment*, *Mutation*, *Gene* and *Amplification Primer*. The figure was obtained with Jambalaya plug-in of Protégé Editor.

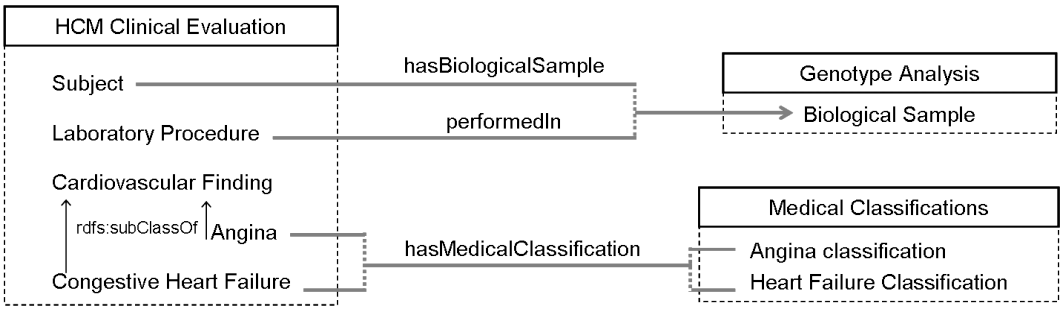

Figure 5. Schematic representation of the non-hierarchical relations through which the three modules that compose the HCM semantic model are connected. Under each module are represented the respective classes that participate in the relations, and between the modules the properties that connect the classes: *hasBiologicalSample*, *performedInBiologicalSample* and *hasMedicalClassification*

Evaluation, with the exception of *Biological Sample*, a class that belongs to the module *Genotype Analysis*. The elements *hasBiologicalSample* and *performedInBiologicalSample* are properties.

Since in the module Genotype Analysis it is possible to connect a Biological Sample with the mutations identified in it, it is through this connection between the modules *HCM Clinical Evaluation* and *Genotype Analysis* that it will be possible to identify each patient's mutations.

In the case of the connection with *Medical Classifications*, it is made between the class *Cardiovascular Finding* (from *HCM Clinical Evaluation*) and any of the classes in the module *Medical Classifications*, through the property *hasMedicalClassification*. There is one such connection for every element in *HCM Clinical Evaluation* that uses a classification defined in *Medical Classifications*. In the actual version of the model, there is one connection for *Angina* and one for *Heart Failure*.

IV. Discussion

The HCM model was initially designed as a single module containing all concepts necessary to characterize a patient in terms of the disease, based on the activities involved in its diagnose. However, this approach presented some difficulties when trying to integrate the molecular biology elements of the biomarker analysis with the clinical elements. While this analysis is considered as an exam by a medical doctor, and so it should be included under the concept *Procedure*, at the same time it has several associated concepts, namely *Gene* and *Mutation*, with information to be maintained, which are not used by the clinician. These different views and characterization needs of the data led to the definition of two modules: one comprising the data elements needed by a medical doctor to evaluate a patient in terms of HCM (*HCM Clinical Evaluation*), and another comprising the data elements needed by a molecular biologist to perform a biomarker analysis in a *Subject* sample (*Genotype Analysis*). In this manner, these data elements are suitably integrated with the clinical evaluation of the disease, and at the same time maintained as laboratory elements that can be managed independently of the *Subject*'s medical data. This separation in modules facilitates not only their individual extension but also their reutilization for different purposes than the characterization of HCM.

The selection of the controlled vocabularies used in the HCM model was based on their content, specifically terms necessary for the description and characterization of a patient in terms of HCM, and on their structural organization. We searched for structures similar to the representation we intended for the semantic model and that better conveyed the vision of the domain experts, what led to the selection of a specific region from two different vocabularies: SNOMED CT and NCIt. However, the adoption of the hierarchical structure of these vocabularies was not always straightforward, namely in the case of the classes *Laboratory Procedure* and *Diagnostic Procedure*. These classes are

currently defined as siblings rather than the first as subclass of the second, even though the procedures considered under *Laboratory Procedure* can, in fact, be considered under *Diagnostic Procedure*(s). We advocate the organization proposed by NCIt, insomuch as it separates procedures that involve the manipulation of a biological sample (*Laboratory Procedure*) from those that do not and are performed directly upon the subject as a whole (*Diagnostic Procedure*).

Still concerning the module *HCM Clinical Evaluation*, its class *Subject* has only two subclasses relevant for the diagnose process: *Patient*, because it instantiates the actual individuals with HCM; and *Family Member*, because it is crucial to know the family history of a patient in relation to HCM manifestations (related both to phenotype and genotype). However, a subclass *Control* was included since it will be necessary for an ulterior identification of correlations between the presence of specific mutations and the *Subject*(s)' physical traits, through the utilization of data mining techniques.

The class *Heart Disease* exists to keep track of other cardiac diseases that either a patient or his family members can suffer from. This information is important for a correct interpretation of the patients' symptoms and exam results.

The HCM semantic model as a whole can be reused and expanded to describe other diseases than HCM, although each module can be used in somewhat different contexts. In this respect, the module *HCM Clinical Evaluation* is the most specific and is best suited for the characterization of heart diseases, while *Genotype Analysis* can be reused/expanded in the context of any disease whose diagnostic or prognostic can be improved by such an analysis. In the case of the module *Medical Classifications*, altough presently containing only two classes representing the classifications used by the consulted medical experts, it can be expanded to include any standard or set of guidelines that refer to the medical aspects of HCM characterization or any other disease.

V. Conclusions

This article presents a novel semantic model that characterizes the diagnose process of the HCM disease. This is a complex disease, both in terms of clinical presentation and number of associated mutations. A translational medicine approach is useful for this type of disease, since it combines genotype and phenotype data in an effort to provide accurate and personalized diagnosis.

The semantic model implements a translational medicine approach to the diagnose of HCM. It was developed using Semantic Web technologies, in particular OWL Lite and controlled vocabularies. During its development, the model evolved into modules, thus facilitating its extension and reutilization.

This model will provide the basic framework for a biomedical system whose purpose is to assist in the integration of the genotype analysis of a patient into his clinical evaluation, in order to improve the HCM diagnose. This will be accomplished through the utilization of data mining techniques that will infer genotype-phenotype correlations, or, more specifically, produce effective models conveying the association of certain mutations with the resulting physical traits. The models thus obtained are expected to be of great interest both in terms of their predictive ability and their practical usability for doctors.

The data to be used in this study will be obtained in a digital format from molecular biology and health care partners such as the *Hospital da Luz*.

Acknowledgments

The authors would like to thank to Nuno Jardim, Leonel Nobrega and Dário Abdulrehman for helpful discussions when building the Activity Map.

This work was supported by the FCT through the Multiannual Funding Program, the doctoral grant SFRH/BD/65257/2009 and the post-doctoral grant SFRH/BPD/20996/2004.

References

[1] M. Tsiknakis, M. Brochhausen, J. Nabrzyski, J. Pucacki, S. G. Sfakianakis, G. Potamias, C. Desmedt, and D. Kafetzopoulos, "A semantic grid infrastructure enabling integrated access and analysis of multilevel biomedical data in support of postgenomic clinical trials on cancer," *IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 12, no. 2, pp. 205–217, Mar. 2008, PMID: 18348950. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18348950

[2] Advancing clinico genomics trials in cancer (ACGT). [Online]. Available: http://www.eu-acgt.org/

[3] cancer Biomedical Informatics Grid (caBIG). [Online]. Available: http://cabig.cancer.gov/

[4] A. Ruttenberg, T. Clark, W. Bug, M. Samwald, O. Bodenreider, H. Chen, D. Doherty, K. Forsberg, Y. Gao, V. Kashyap, J. Kinoshita, J. Luciano, M. S. Marshall, C. Ogbuji, J. Rees, S. Stephens, G. Wong, E. Wu, D. Zaccagnini, T. Hongsermeier, E. Neumann, I. Herman, and Cheung K-H, "Advancing translational research with the semantic web," *BMC Bioinformatics*, vol. 8, p. S2, 2007.

[5] V. Kashyap, "From the bench to the bedside: the role of Semantic Web and Translational Medicine for enabling the next generation healthcare enterprise," in *Biomedical Engineering Systems and Technologies*, A. Fred, J. Filipe, and H. Gamboa, Eds. New York: Springer, 2008, pp. 35–56, Communications in Computer and Information Science.

[6] W3C semantic web standards. [Online]. Available: http://www.w3.org/standards/semanticweb/

[7] Systematized nomenclature of medicine-clinical terms (SNOMED). [Online]. Available: http://www.ihtsdo.org/snomed-ct/

[8] N. Sioutos, S. Coronado, M. W. Haber, F. W. Hartel, W. L. Shaiu, and L. W. Wright, "NCI thesaurus: A semantic model integrating cancer-related clinical and molecular information," *J Biomed Inform*, vol. 40, pp. 30–43, 2007.

[9] The ontology of clinical research (OCRe). [Online]. Available: http://rctbank.ucsf.edu/home/ocre.html

[10] R. C. Gudivada, X. A. Qu, J. Chen, A. G. Jegga, E. K. Neumann, and B. J. Aronow, "Identifying disease-causal genes using semantic web-based representation of integrated genomic and phenomic knowledge," *Journal of Biomedical Informatics*, vol. 41, no. 5, pp. 717–729, Oct. 2008, PMID: 18755295. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18755295

[11] B. J. Maron, M. S. Maron, E. D. Wigle, and E. Braunwald, "The 50-year history,controversy, and clinical implications of left ventricular outflow tract obstruction in hypertrophic cardiomyopathy: from idiopathic hypertrophic subaortic stenosis to hypertrophic cardiomyopathy," *J Am Coll Cardiol*, vol. 54, pp. 191–200, 2009.

[12] R. Alcalai, J. G. Seidman, and C. E. Seidman, "Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics," *J Cardiovasc Electrophysiol*, vol. 19, pp. 104–110, 2008.

[13] Human genome mutation database. [Online]. Available: http://www.hgmd.org

[14] Harvard sarcomere mutation database. [Online]. Available: http://genepath.med.harvard.edu/~seidman/cg3/

[15] N. Cardim, A. Perrot, S. Santos, P. Morgado, M. Pádua, S. Ferreira, R. P. Reis, C. Monteiro, T. Ferreira, and J. M. Correia, "Hypertrophic cardiomyopathy in a portuguese population: mutations in the myosin-binding protein c gene," *Rev Port Cardiol*, vol. 24, pp. 1463–1476, 2005.

[16] D. Brito, P. Richard, R. Isnard, J. Pipa, M. Komajda, and H. Madeira, "Familial hypertrophic cardiomyopathy: the same mutation, different prognosis. comparison of two families with a long follow-up." *Rev Port Cardiol*, vol. 22, pp. 1445–1461, 2003.

[17] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Knowledge Systems, AI Laboratory, Stanford University, Tech. Rep. KSL-01-05, 2001.

[18] L. L. Constantine, "Interaction design and model-driven development," in *Model Driven Engineering Languages and Systems*, ser. Lecture Notes in Computer Science. Springer, 2009, p. 377.

[19] N. F. Noy, N. H. Shah, P. L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D. L. Rubin, Storey M-A, C. G. Chute, and M. A. Musen, "BioPortal: ontologies and integrated data resources at the click of a mouse," *Nucleic Acids Res*, vol. 37, pp. W170–W173, 2009.

[20] National center for biomedical ontology bioportal. [Online]. Available: http://bioportal.bioontology.org

[21] Protégé ontology editor. [Online]. Available: http://protege.stanford.edu

[22] Canadian cardiovascular society. [Online]. Available: http://www.ccs.ca/home/index_e.aspx

[23] *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*, Criteria Committee of the New York Heart Association, Boston, 1994.