

**Bernardo Gonçalves**

**An Ontological Theory of the Electrocardiogram with  
Applications**

Vitória - ES, Brazil

May 13, 2009

**Bernardo Gonçalves**

# **An Ontological Theory of the Electrocardiogram with Applications**

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo para obtenção do título de Mestre em Informática.

Orientador:

**José Gonçalves Pereira Filho**

Co-orientador:

**Giancarlo Guizzardi**

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA  
DEPARTAMENTO DE INFORMÁTICA  
CENTRO TECNOLÓGICO  
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Vitória - ES, Brazil

May 13, 2009

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

---

G635o Gonçalves, Bernardo, 1982-  
An ontological theory of the electrocardiogram with  
applications / Bernardo Gonçalves. – 2009.  
150 f. : il.

Orientador: José Gonçalves Pereira Filho.

Co-Orientador: Giancarlo Guizzardi.

Dissertação (mestrado) – Universidade Federal do Espírito Santo,  
Centro Tecnológico.

1. Ontologia. 2. Informática na medicina. 3. Modelagem de  
dados. 4. Inteligência artificial. I. Pereira Filho, José Gonçalves.  
II. Guizzardi, Giancarlo. III. Universidade Federal do Espírito  
Santo. Centro Tecnológico. IV. Título.

CDU: 004

---

Dissertação de Mestrado sob o título “*An Ontological Theory of the Electrocardiogram with Applications*”, defendida por Bernardo Gonçalves em May 13, 2009, em Vitória, Estado do Espírito Santo, e aprovada por unanimidade pela banca examinadora constituída pelos doutores:

---

Prof. Dr. José Gonçalves Pereira Filho  
Departamento de Informática - UFES  
Orientador

---

Prof. Dr. Giancarlo Guizzardi  
Departamento de Informática - UFES  
Co-orientador

---

Prof. Dr. João Paulo Almeida  
Departamento de Informática - UFES  
Membro interno

---

Prof. Dr. Frederico Fonseca  
College of IST - Pennsylvania State University  
Membro externo

# Abstract

The fields of Medical- and Bio-informatics are bearing witness of the application of the discipline of Formal Ontology to the representation of biomedical entities and (re-)organization of medical terminologies also in view of advancing electronic health records (EHR). In this context, the electrocardiogram (ECG) defines one of the prominent kinds of biomedical data. As a vital sign, it is an important piece in the composition of the EHR of today, as likely in the EHR of the future. This thesis introduces an ontological analysis of the ECG grounded in the Unified Foundational Ontology (UFO) and axiomatized in First-Order Logic (FOL). With the goal of investigating the phenomena underlying this cardiological exam, we deal with the sub-domains of human heart electrophysiology and anatomy. We then outline an ECG ontology meant to represent *what the ECG is* on both sides of the patient and of the physician. The ontology is implemented in the semantic web technology OWL with its SWRL extension.

The ECG Ontology makes use of basic relations standardized in the OBO Relation Ontology for the biomedical domain. In addition, it takes inspiration in the Foundational Model of Anatomy (FMA) and applies the Ontology of Functions (OF). Besides the ECG ontological theory itself, two applications of the ECG Ontology are also presented here. The first one is concerned with the off-line integration of ECG data standards, a relevant endeavor for the progress of Medical Informatics. The second one in turn comprises a reasoning-based web system that can be used to offer support for interactive learning in electrocardiography / heart electrophysiology. Overall, we also reflect on the ECG Ontology as well as on its two applications to provide evidence for benefits achieved with the employment of methodological principles - in terms of both ontological foundations and ontology engineering - in building a domain ontology.

# Dedictory

To my mother.

# Acknowledgements

The Earth has made about 800 rotations around itself since I started my Master's in Informatics. It is inestimable for me how much I have been growing in terms of scientific maturity from then on. First of all, I would like to thank my family and my dear girlfriend Marcelle Olivier for all the support they provided me with along that time. Without their comprehension and encouragement, the work reported in this thesis would not be possible. I am also grateful to my English teacher Sarah for her pleasant dedication in the improvement of my English.

In the context of UFES, the first person for whom I owe this opportunity and would like to express my gratitude is prof. Rosane Caruso. She has been the first person who believed me as a student and applicant for scientific initiation. Furthermore, her passion for Logics has inspired plenty of students at UFES/DI, and me in particular. Thanks a lot, Rô!

The next person who I am indebted to is prof. José Gonçalves Pereira Filho. He is responsible for over my two first years of scientific initiation. Although our discussions were unfortunately always with time to end due to his uncountable number of important duties, they were always funny, fruitful and inspiring. He dedicated his time on the task of making me a technical writer. Zé Gonçalves is also an admirable person due to his belief in and fight for making Brazil, in general, and Espírito Santo, in particular, a nice place in terms of higher education. But among many other good things, the most important one that comes to my mind when I think of Zé Gonçalves, is that in the most difficult situations I was in, he was there for me. Thank you very much for all, Zé!

Now it is time to put prof. Giancarlo Guizzardi in the story. Since Giancarlo came to UFES in 2006 my vision of research and Computer Science has been greatly expanded. The very first lecture I took from him was for me such a breakthrough that sometimes I find myself grateful to life for putting me in the right place at the right time. Since then Giancarlo has been for me a kind of "best partner", to cite an expression brought by Renata Guizzardi. This partnership means for me somewhat I see in reading Ancient Philosophy, like in the ancient initiation of young students in Philosophy. If I would have to choose one aspect of Giancarlo's guidance which has been the most fundamental for me, I would say that it is the very balance between rationality and passion in seeking the truth. For three years he has been encouraging me in doing research with scientific impartiality, but also with a pre-socratic-like enthusiasm. Gian, I have no words to express my gratitude for all that you have been teaching me, even implicitly. Thank you for reading every page of this thesis; but above all, thank you for showing me the way to become an ontologist.

I would also like to thank prof. Berilhes Garcia for being gently open for a number of coffee breaks which the subject of discussion somehow insisted to fall into Philosophy of Science. I'm lucky as well to have had the opportunity to do this Master's course besides Veruska Zamborlini, Raphael Santos and Felipe Frechiani. Either in technical discussions or in coffee breaks and happy hours, they have been essential parts of my master's trajectory. Veruska has been a great research colleague, who contributed significantly to the work reported on this thesis. Raphael and Felipe in turn, besides our good discussions about research, are mostly my "programming partners", always there for sharing nice programming tips. Finally, I would like to thank my classmates from the graduation

in Computer Science at UFES; my colleagues André Costa and Luiz Rodrigo from the TeleCardio project; William Hisatugu for being a very nice lab colleague; and also all professors from whom I took lectures at UFES/DI.



# Table of Contents

## List of Figures

## List of Tables

<b>1 Introduction</b>	p. 15
1.1 Motivation . . . . .	p. 15
1.2 Goals and Scope . . . . .	p. 16
1.2.1 Goals . . . . .	p. 16
1.2.2 Research questions . . . . .	p. 18
1.3 Approach and Structure . . . . .	p. 18
<b>2 Background, Part I: Ontology in Computer Science</b>	p. 20
2.1 Ontology, Ontology and Ontology . . . . .	p. 20
2.1.1 The Beginning . . . . .	p. 20
2.1.2 Ontology Definition . . . . .	p. 22
2.1.3 How to Talk about Good Ontologies? . . . . .	p. 22
2.2 Ontological Foundations . . . . .	p. 23
2.2.1 Formal Ontological Principles . . . . .	p. 23
2.2.2 Top-Level Ontologies . . . . .	p. 24
2.3 Ontology Formalisms . . . . .	p. 25
2.4 Ontology Engineering . . . . .	p. 28
2.5 Conclusions . . . . .	p. 29
<b>3 Background, Part II: Biomedical Ontology</b>	p. 30
3.1 Biomedical Terminologies and Ontologies . . . . .	p. 30
3.1.1 UMLS Semantic Network . . . . .	p. 31
3.1.2 NCI Thesaurus . . . . .	p. 32

3.1.3	SNOMED-CT . . . . .	p. 33
3.1.4	Gene Ontology . . . . .	p. 35
3.1.5	Foundational Model of Anatomy . . . . .	p. 38
3.2	Biomedical Ontologies' Adherence to Ontology . . . . .	p. 40
3.3	Concept- vs. Realism-orientation . . . . .	p. 42
3.4	Applications of Biomedical Ontologies . . . . .	p. 44
3.5	Conclusions . . . . .	p. 45
<b>4</b>	<b>Materials &amp; Methods</b>	p. 46
4.1	Materials . . . . .	p. 46
4.1.1	Unified Foundational Ontology and OntoUML . . . . .	p. 46
4.1.2	OBO Relation Ontology . . . . .	p. 50
4.1.3	Ontology of Functions . . . . .	p. 51
4.1.4	OWL DL / SWRL . . . . .	p. 53
4.2	Methods . . . . .	p. 55
<b>5</b>	<b>The ECG Ontological Theory</b>	p. 57
5.1	Preliminaries: Conventions & Epistemological Assumptions . . . . .	p. 57
5.2	Anatomy for the ECG . . . . .	p. 59
5.3	Heart Electrophysiology . . . . .	p. 68
5.4	The Electrocardiogram . . . . .	p. 75
5.5	Basic ECG Interpretation . . . . .	p. 83
5.6	From the ECG to Heart Electrophysiology . . . . .	p. 85
5.7	An ECG Ontology . . . . .	p. 88
5.7.1	Competence Questions . . . . .	p. 88
5.7.2	Documentation . . . . .	p. 90
5.8	Conclusions . . . . .	p. 98
<b>6</b>	<b>ECG Ontology Implementation</b>	p. 100
6.1	Basic Design Patterns . . . . .	p. 100
6.2	The ECG OWL Ontology . . . . .	p. 101
6.2.1	The OBO RO Extension . . . . .	p. 102

6.2.2	Anatomy OWL Sub-Ontology . . . . .	p. 103
6.2.3	Physiology OWL Sub-Ontology . . . . .	p. 104
6.2.4	ECG OWL Sub-Ontology . . . . .	p. 105
6.2.5	FOL Formulae as OWL Restrictions and SWRL Rules . . . . .	p. 106
6.3	ECG Ontology Evaluation . . . . .	p. 109
6.4	Discussion . . . . .	p. 110
6.5	Conclusions . . . . .	p. 111
<b>7</b>	<b>Application in Conceptual Modeling</b>	p. 113
7.1	ECG Data Standardization: An Ongoing Story . . . . .	p. 113
7.2	The Reference ECG Data Formats . . . . .	p. 115
7.2.1	AHA/MIT-BIH . . . . .	p. 116
7.2.2	SCP-ECG . . . . .	p. 117
7.2.3	FDA XML / HL7 aECG . . . . .	p. 119
7.2.4	Discussion . . . . .	p. 123
7.3	Ontology for Semantic Interoperability . . . . .	p. 123
7.4	An Integration Experiment . . . . .	p. 125
7.5	Conclusions . . . . .	p. 127
<b>8</b>	<b>Application in Symbolic AI</b>	p. 129
8.1	ECG Data Input: The QT Database from Physionet . . . . .	p. 130
8.2	Application Technologies and Architectural Overview . . . . .	p. 130
8.3	Application Logic . . . . .	p. 131
8.3.1	ECG Chart Service . . . . .	p. 132
8.3.2	Inference Service . . . . .	p. 132
8.3.3	Flash Media Objects . . . . .	p. 133
8.4	Performance Evaluation . . . . .	p. 133
8.5	Discussion . . . . .	p. 134
8.5.1	Reasoning over Universals and Particulars . . . . .	p. 134
8.5.2	Educational Animations . . . . .	p. 135
8.6	Conclusions . . . . .	p. 136

**9 Discussion & Final Considerations** p. 137

9.1 Revisiting our Goals and Research Questions . . . . . p. 137

9.2 Significance . . . . . p. 138

9.3 Limitations . . . . . p. 139

9.4 Open Problems and Future Work . . . . . p. 140

9.5 Final Considerations . . . . . p. 140

**References** p. 142

# List of Figures

1	Gross subject areas for medical ontologies . . . . .	p. 17
2	Overview of the thesis structure . . . . .	p. 19
3	Tree of Porphyry, with Aristotle's categories and their differentiae . . . . .	p. 21
4	The intended models of a logical language . . . . .	p. 23
5	Three different possible models for representing the concept of customer . . . . .	p. 27
6	Portion of the UMLS semantic network - source: (1). . . . .	p. 32
7	A NCI thesaurus' query result on the term 'tumor-derived' (2). . . . .	p. 33
8	SNOMED-CT's tree view . . . . .	p. 34
9	The three sub-ontologies of the Gene Ontology (3). . . . .	p. 37
10	A GO's query result on the term 'biological process' (3). . . . .	p. 37
11	The Foundational Model of Anatomy . . . . .	p. 39
12	Excerpt of the UFO ontology . . . . .	p. 47
13	Two exemplary models employing OF (source: (4)). . . . .	p. 53
14	The OWL Classes view . . . . .	p. 54
15	Protégé also supports editing rule bases . . . . .	p. 55
16	Anatomical entity and its partition into material and immaterial entities . . . . .	p. 59
17	The <i>is-a</i> taxonomy descending from Organ component . . . . .	p. 60
18	Internal anatomy of the wall of the heart . . . . .	p. 61
19	The <i>is-a</i> taxonomy descending from Region of organ component . . . . .	p. 62
20	The <i>is-a</i> taxonomy descending from Portion of tissue . . . . .	p. 62
21	Referred subdivisions of the conducting system of the heart . . . . .	p. 62
22	The conducting system of the heart . . . . .	p. 63
23	Partonomy of anatomical entities which concern the ECG . . . . .	p. 65
24	Material anatomical categories Anatomical cluster and Portion of body substance . . . . .	p. 66
25	Relations involving myocytes of subdivisions of the heart conducting system . . . . .	p. 67

26	Relations involving the <b>Body surface</b> . . . . .	p. 68
27	Propagation of the cardiac electrical impulse . . . . .	p. 69
28	Two disjoint phases of myocytes . . . . .	p. 70
29	Cardiac circulation . . . . .	p. 70
30	Function To generate CEI represented in the OF framework . . . . .	p. 71
31	Function To conduct CEI (in two different manifestations) represented in the OF framework . . . . .	p. 72
32	Function To restore EPs represented in the OF framework . . . . .	p. 72
33	Model of heart electrophysiological functions. . . . .	p. 73
34	Model of the ECG recording session context . . . . .	p. 76
35	Model of the ECG acquisition mechanism . . . . .	p. 77
36	ECG leads . . . . .	p. 79
37	A typical cycle in the ECG waveform . . . . .	p. 80
38	Model of the ECG waveform (on the side of the physician) . . . . .	p. 81
39	Mapping relations between ECG forms and electrophysiological processes . . . . .	p. 84
40	Import relationships of the ECG Ontology . . . . .	p. 88
41	General picture of the ECG OWL Ontology edited in Protege . . . . .	p. 102
42	Tree-based data model of the AHA/MIT-BIH . . . . .	p. 117
43	Conceptual model of the AHA/MIT-BIH . . . . .	p. 117
44	Tree-based data model of the SCP-ECG . . . . .	p. 119
45	Conceptual model of the SCP-ECG . . . . .	p. 120
46	Tree-based data model of the FDA XML / HL7 aECG . . . . .	p. 122
47	Conceptual model of the FDA XML / HL7 aECG . . . . .	p. 122
48	Integration between the AHA/MIT-BIH conceptual model and the ECG Ontology . . . . .	p. 125
49	Integration between the SCP-ECG conceptual model and the ECG Ontology . . . . .	p. 126
50	Integration between the FDA XML / HL7 aECG conceptual model and the ECG Ontology . . . . .	p. 127
51	Screenshot of the reasoning-based web application . . . . .	p. 130
52	Application architectural overview . . . . .	p. 131

# List of Tables

1	The relations of the OBO Relation Ontology that we make use in this thesis. . . . .	p. 51
3	ECG Ontology class dictionary: sub-ontology of anatomy . . . . .	p. 90
2	ECG Ontology relations and their meta-properties . . . . .	p. 95
4	ECG Ontology class dictionary: sub-ontology of Heart Electrophysiology . . . . .	p. 96
5	ECG Ontology class dictionary: ECG ontology . . . . .	p. 97
6	OWL object properties derived from ECG Ontology's relations and their features . . . . .	p. 103
7	Evaluation results for the ECG Ontology implementation . . . . .	p. 110
8	Sections of a SCP-ECG record and their descriptions. . . . .	p. 118
9	Correspondence relations between classes in the ECG Ont. and the ECG standards . . . . .	p. 126
10	Timing measurements (in ms) for (I1) . . . . .	p. 134
11	Timing measurements (in ms) for (I3) . . . . .	p. 134
12	Timing measurements (in ms) for (I2) . . . . .	p. 135

# 1 Introduction

This master thesis presents our research on ontology development in the field of Bioinformatics. It contributes to the field of biomedical ontology with an ontological theory of the electrocardiogram and its application. The thesis focuses on technical aspects, but touches upon less technical aspects as well.

First of all, this chapter introduces our motivation (Section 1.1) for the research. It then defines our goals and scope (Section 1.2), which are supported by research questions this thesis is meant to answer. Subsequently, the chapter discusses the approach employed for the research and depicts the thesis structure (Section 1.3).

## 1.1 Motivation

The fields of Medical- and Bio-informatics have seen in recent years growing research efforts regarding the representation of biological entities, e.g. (5), and (re-)organization of medical terminologies and EHR - electronic health records, e.g. (6). The motivation is (basically) to set the ground for: (i) biologists and physicians to store and communicate biomedical information and patient-related data effectively; and (ii) gradually integrating these sources in the development of next generation knowledge-based biomedical computer applications. These applications are meant to provide support in basic science and clinical research, as well as in the delivery of more efficient health care services. As posed by Rosse and Mejino Jr (7), *“such a widening focus in bioinformatics is inevitable in the post-genomic era, and the process has in fact already begun”*.

However, in spite of these broad perspectives, there still exist a number of problems and challenges to overcome. The more biological and medical knowledge presented in scientific papers increases and becomes mutually dependent, the more complex is the task of representing it while keeping a consistent integration. Besides, patient data has been increasingly stored digitally in EHR's as a result of the growing use of information systems in health environments. Therefore, the need for structuring this vast amount of existing biomedical knowledge and data grows in the same pace (8). This is a fundamental need not only to ease an effective data access as usual, but also to afford formal analysis for further use in problem-solving and for developing and testing hypotheses (9). To get further into the discussion, let us take into consideration this “simple” interoperability-problem example below given by James Cimino (10, p. 394) in that salient paper conveying desiderata for controlled medical vocabularies a decade ago.

Consider, for example, how a computer-based medical record system might work with a diagnostic expert system to improve patient care. In order to achieve optimal integration of the two, transfer of patient information from the record to the expert would need to be automated. In one attempt to do so, the differences between the controlled vocabularies of the two systems was found to be the major obstacle - even when both systems were created by the same developers.

One might argue that the advances reached since then have not been so extensive. As a matter of fact,



*interoperability* is still a challenge to cope with *even when both systems were created by the same developers*.

Nevertheless, especially after the Semantic Web envision (11), the term ‘ontology’ has appeared as a solution (and occasionally even proclaimed as the ultimate solution) for all these issues. Indeed, *ontology* has been promoted as a technique to build advanced information systems (12), for which Biomedicine is a rich field of application. In a survey article (13), Bodenreider and Stevens discuss the influence ontologies have been impinged in Bioinformatics. It is nowadays such that there has been a shift from a strictly technology-oriented paradigm to a philosophically founded one. There is an extensive list of current research initiatives promoting the ontology-based approach to handle the representation of (subdomains of) the biomedical domain, e.g. (14, 15, 8, 16). A prominent initiative for gathering biomedical ontologies in a principled way is the Open Biomedical Ontologies (OBO) foundry (17). Up to this point, it comprises over 60 ontologies each of which, although varying a lot in terms of granularity, canonicity and developmental stage, aims at representing a clearly bounded subject-matter.

Among the most referred ontologies in OBO, one might cite the Foundational Model of Anatomy (FMA) (7), the Gene Ontology (GO) (18), and the Chemical Entities of Biological Interest (ChEBI) (19). While the FMA deals with the structure of the mammalian (especially the human body), GO covers attributes of gene products in all organisms and ChEBI targets molecular entities which are products of nature or synthetic products used to intervene in the processes of living organisms. However, despite the fact that the domain of human heart electrophysiology is of significant interest in Biomedicine, an ontology of heart electrophysiology is still missing in OBO as well as in the biomedical ontology literature<sup>1</sup>. Furthermore, although the electrocardiogram (ECG) defines one of the prominent kind of biomedical data, as far as we know, it has not yet been addressed in the biomedical ontology literature. Nonetheless, the ECG appears in an outline of the gross subject areas for medical ontologies provided by Bodenreider and Stevens in their survey article aforementioned (13), see Figure 1.

The ECG is the most frequently applied test for measuring heart activity in Cardiology (22). In recent years, both the storage and transmission of ECG records have been object of standardization initiatives. Among the foremost ECG standards, one might refer to SCP-ECG (23) and FDA XML (24) / HL7 aECG (25). However, the focus of such standards is mostly on how data and information should be represented in computer and messaging systems (17, p. 1252), (26, p. 254). On the other hand, there is a need for concentrating on the proper representation of the biomedical reality under scrutiny (27, 28). Namely, on *what the ECG is*, on both sides of the patient and of the physician. This is clearly relevant, since the ECG, as a vital sign, is an important piece in the composition of the EHR of today, as likely in the EHR of the future.

## 1.2 Goals and Scope

In the light of that motivation, this section describes our goals and scope.

### 1.2.1 Goals

In line with the biomedical ontology literature, this thesis is intended to represent a clearly bounded subject-matter in Biomedicine. The target domain (or universe of discourse) here is the ECG, which is dealt with as the subject of ontological analysis. Our main goal is defined as follows:

---

<sup>1</sup>We are aware of two ongoing research initiatives which fall roughly in heart electrophysiology. Rubin et al. (20) present a symbolic, ontologically-guided methodology for representing a physiological model of the circulation as an alternative to mathematical models commonly employed. In turn, Cook et al. (21) are putting effort in an extension of the FMA to cover physiology.

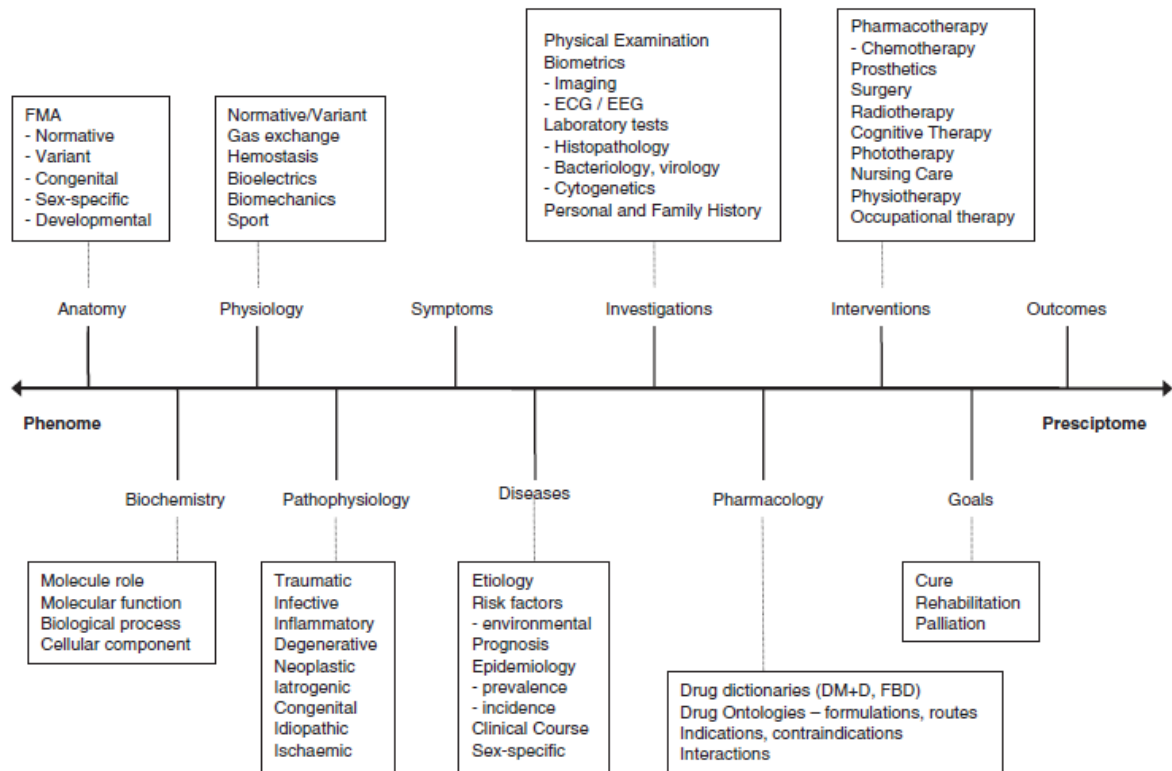


Figure 1: Gross subject areas for medical ontologies arranged in a space from the *phenome* (space of observable characteristics) to the *prescriptome* (space of treatments). The ECG lies in the “investigations” class, on the middle. Source: (13).

*“To develop an ontological theory of the ECG (independent of application and codification language), and further apply it by providing evidence of its benefits”*

By reaching this main goal, we aim to contribute to the biomedical ontology literature. The first result expected is then an ontology of ECG, which is named in this text **ECG Ontology**. The second result to be achieved is a twofold application of the ECG Ontology: firstly, in the context of Conceptual Modeling (CM), the ontology is used to foster interoperability of ECG standards; secondly, in the Artificial Intelligence (AI) context, the ontology is used in a reasoning-based application. The main goal can then be refined into the following specific goals.

1. We aim at developing two ontology artifacts (cf. Section 2.4): (i) an ontologically well-founded theory of the subject domain meant to be strongly axiomatized for constraining as much as possible its intended meaning; and (ii) a computable artifact derived from that theory for automated reasoning and information retrieval. The former is referred to further on in this text as *ontological theory* or *reference conceptual model*, while the latter is referred to as *ontology implementation* or *ontology codification*.
2. Provide evidence for the following hypothesis: an ECG reference ontology can be used to foster interoperability of different conceptual models in the ECG domain.
3. Likewise, provide evidence for the assumption that an ECG ontology implementation derived from its reference counterpart can be used with genuine benefits in a reasoning-based computer application.

By reaching the specific goals above we expect to contribute to the ontology engineering literature as well.

**Non-Goal**

It is not a goal of this thesis to model the ECG domain in a quantitative approach (say, with mathematical equations). There are solid works in this direction, see e.g. (29). In this thesis we rather address a qualitative - somewhat *naïve physics* (30) - modeling of the ECG domain. We then aim here to provide a qualitative counterpart to Geselowitz's article "On the Theory of the Electrocardiogram" (22).

**1.2.2 Research questions**

To reach our goals, we pursue the following research questions:

1. What is the ECG in essence?
2. What can an off-line ECG ontological theory (or reference conceptual model) be used for?
3. Is it worthwhile to derive an ontology implementation from an ontological theory?
4. What can be done by using the codification of an ECG ontology in a reasoning-based computer application?  
Are there any benefits, say, when compared to other AI formalisms? Which are them?

Answering these questions constitutes a pre-requisite to reach our goals satisfactorily. They thus are revisited for discussion in Section 9.1.

**1.3 Approach and Structure**

We have employed an iterative approach along the development of this research. That is, we have worked towards a first version of the ECG Ontology, implemented, evaluated and applied it at a first glance, and looped back to this cycle a second time. This development cycle assumes an ontology engineering approach that, analogous to any other engineering process, comprises the phases of analysis, design and implementation, followed by an evaluation.

The structure of this thesis reflects the goals we have been pursuing throughout the research. First of all, however, we provide a theoretical background of (i) Ontology in Computer Science (Chapter 2), and (ii) how ontologies have been developed and applied in Biomedicine (Chapter 3). Finally, we present our methodological choices and the materials used in our research (Chapter 4). We then proceed to introduce the ECG ontological theory proposed in this thesis (Chapter 5). Its implementation is presented subsequently in Chapter 6. These two chapters are meant to reach Goal 1. As follows we present in Chapter 7 an application of the ECG theory in Conceptual Modeling to support interoperability of ECG standards. This chapter is meant to reach Goal 2. Moreover, we present in Chapter 8 an application in Symbolic AI of the ECG ontology implementation. This chapter is to accomplish Goal 3. Finally, we conclude the thesis (Chapter 9) with a discussion on its contributions and significance; but also by referring to its limitations and future work; we then provide our final considerations. In summary, the thesis' structure and its connection to the goals mentioned above is synthesized in Figure 2.

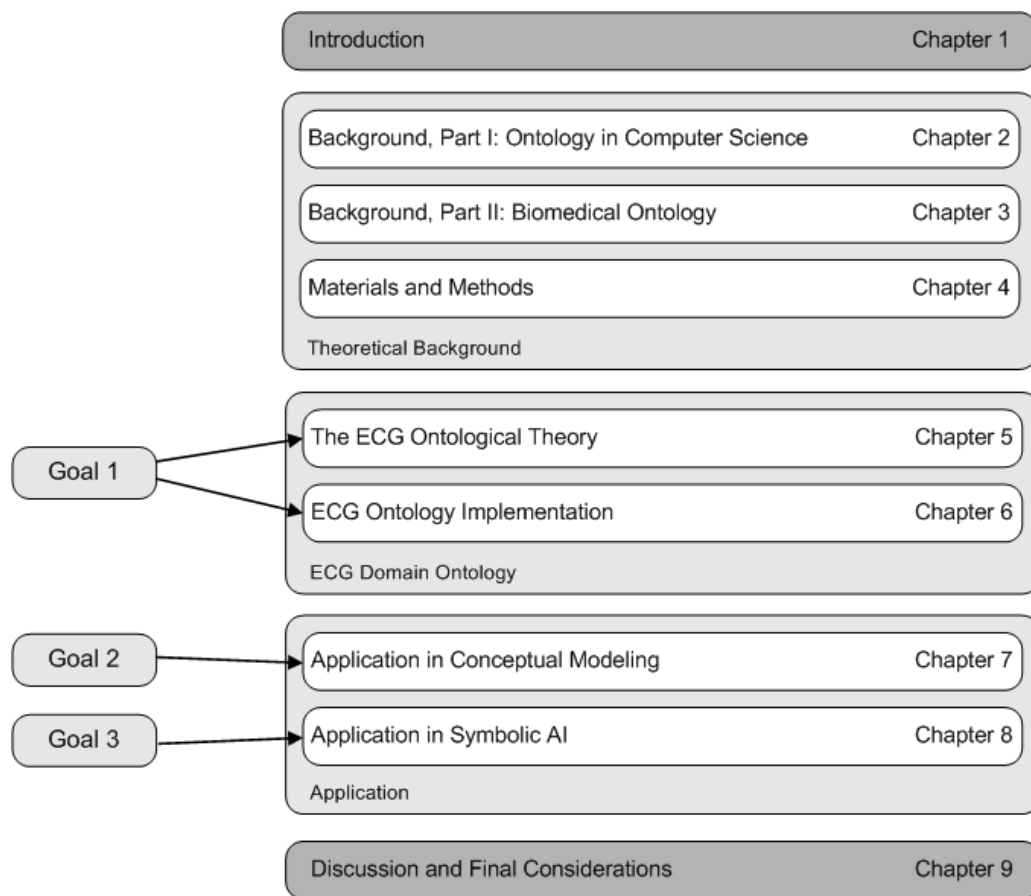


Figure 2: Overview of the thesis structure relating the goals with the chapters in which they are accomplished.

## 2 Background, Part I: Ontology in Computer Science

This chapter is devoted to provide a brief background of *ontology* in Computer Science. Our aim here is not to give a deep account of it, but rather to introduce relevant aspects and issues which are referred to further on in this text. For a basic reading, we suggest (12), (31) or (32), and (33) or (34, Chapter 3). We start the chapter with some historical considerations, and proceed by providing referred definitions, philosophical and methodological issues that underlie the theme of ontology. The chapter is then concluded with a summary of the key points developed throughout it.

### 2.1 Ontology, Ontology and Ontology

#### 2.1.1 The Beginning

The systematic study of Metaphysics has been addressed in Western Philosophy at least since Aristotle. A curious thing about it is that the Aristotle's endeavor of representing the structure of reality apparently was primarily put on the biological domain. Biology seems to be in fact a classical domain of ontological application, for that the most likely first ontologist is in general recognized as the “father of Biology”. As shown in Figure 3, Aristotle's categories were arranged by *genus* (supertype) and *species* (subtype). The specific task of distinguishing species under the same genus was made use of *differentiae*, some sort of properties that distinguish a given species from another under the same genus. For instance, *material* and *immaterial* are both *differentiae* to distinguish the species *body* and *spirit* under the genus *substance*. This can arguably be said the first formal ontological principle used to support ontological decisions.

Over the years, the subfield of Philosophy that came to be called *Analytical Metaphysics* has accumulated a significant body of analytical tools for ontological problems. Formal principles of classification have been elaborated, and many of them rest already on a wide consensus among philosophers. In Computer Science (CS), on the other hand, the term ‘ontology’ has been used sometimes with a weak (if at all) connection to the philosophical discipline of *Formal Ontology*. In the following, we briefly discuss the multiple meanings computer scientists have been assigning to ontology.

As discussed by Giancarlo Guizzardi (33, p. 19), the term ‘ontology’ in the computer and information science literature appeared for the first time in 1967 (36), in a work on the foundations of data modeling by Mealy, inspired by his reading of Quine (37). Barry Smith and Chris Welty in turn draw attention to another early use of the term by John McCarthy in the context of AI also influenced by Quine, cf. (31, p. v). Patrick Hayes, in his “Naive physics I: Ontology for liquids” (38), and John Sowa in his “Conceptual structures” (39) have subsequently referred to



*Supreme genus:*

*Differentiae:*

*Subordinate genera:*

*Differentiae:*

*Subordinate genera:*

*Differentiae:*

*Proximate genera:*

*Differentiae:*

*Species:*

*Individuals:*

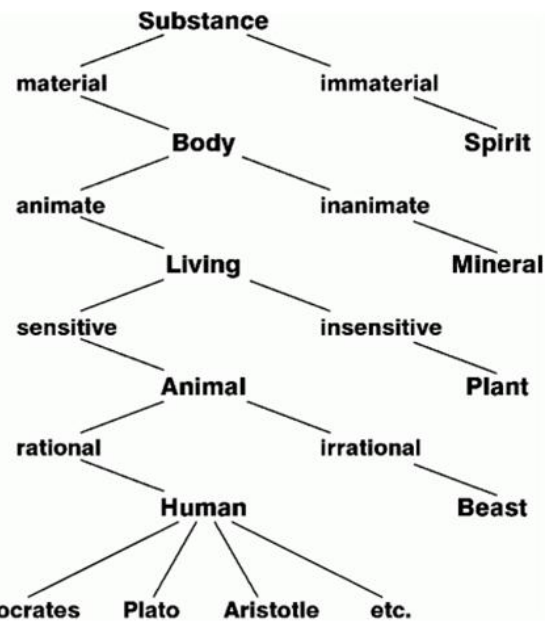


Figure 3: Tree of Porphyry, with Aristotle's categories and their differentiae. Lines represent is-a (subsumption) relationships between categories (source: Philosophy of Aristotle's homepage at University of Washington). The tree of Porphyry can be found also in (35).

'ontology' as well. The story is told by Smith and Welty (31) in the following terms. Initially, symbolic AI has been focused on the development of systems that "know", so-called *expert systems*. They were meant to simulate knowledge through the use of automated reasoning mechanisms. However, as these mechanisms became more standardized over time, the theories expressed in such intelligent systems (so-called knowledge bases) became a focus of attention as well, and the field of knowledge engineering was born (40). Meanwhile, two other fields of CS, namely Database Systems and Software Engineering, started to recognize the need for advancing *conceptual modeling* techniques (31, p. iv). In the latter, on one side, ontology development has been taken as a means for *domain modeling*. This is meant to promote reusable conceptual models capable of facing the increase of size and complexity of software. In the former, on the other side, ontology is seen as a means to foster consistent database conceptual modeling in view of further interoperability with heterogeneous information systems.

Nonetheless, in spite of the Quine's influence, as the significance of the term grew in CS, the ambiguity was increased in the same pace. We then conclude this subsection by putting it literally from Smith and Welty (31, p. v).

Despite encouragement from these influential figures, most of AI chose not to consider the work of the much older overlapping field of philosophical ontology, preferring instead to use the term 'ontology' as an exotic name for what they'd been doing all along knowledge engineering. This resulted in an unfortunate skewing of the meaning of the term as used in the AI and information systems fields, as work under the heading of 'ontology' was brought closer to logical theory, and especially to logical semantics, and it became correspondingly more remote from anything which might stand in a direct relation to existence or reality. Some may argue that this meaning is appropriate for a computer system, as a logico-semantic theory will, in fact, define the kinds and structures of objects, properties, events, processes and relations *that exist in the system*. On the other hand, many are now arguing that the very lack of grounding in external reality is precisely what created the problems, so pressing for the information industry today, of legacy system integration. How can we make older systems with different conceptual models but overlapping semantics work together, if not by referring to the common world to which they all relate?

### 2.1.2 Ontology Definition

In that quite chaotic context, the first attempt to come with an ontology definition in CS has been made by Thomas Gruber in 1995. His definition, which up to these days is the most referenced one, states that an ontology is an *“explicit specification of a conceptualization”* (41, p. 907). It however, requires some word for what exactly Gruber calls *conceptualization*. In his view, a *conceptualization* is *“the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.”* (41, p. 907). Gruber’s definition seems to be the most popular one. In the formal ontology literature, however, it has been criticized on the grounds that it allows a broad interpretation (12, p. 5), (31, p. vi).

In face of this, Nicola Guarino has attempted to formalize a more elaborate ontology definition in view of clarifying the term confusion and assigning an intensional account to the notion of conceptualization. First, beyond ‘Ontology’ with the capital ‘O’ meaning the philosophical discipline, he refers to ‘ontology’ in the philosophical sense as *“a particular system of categories accounting for a certain vision of the world. As such, this system does not depend on a particular language: Aristotle’s ontology is always the same, independently of the language used to describe it”* (12, p. 4). On the side of CS, rather, Guarino (12, p. 4) refers to ‘ontology’ as *“an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words”*. He adds still that this set of assumptions is mostly expressed by means of a first-order logical theory, with vocabulary words appearing as unary or binary predicate names, respectively called concepts and relations. A hierarchy of concepts related by subsumption relationships can be said to be the most simple form of ontology. It, however, becomes more elaborate if suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation.

In an attempt to get over with the terminological impasse, Guarino has chosen to keep using ‘ontology’ only in the CS reading. A new term, viz., ‘conceptualization’, has been assigned to the philosophical reading, such that *“two ontologies can be different in the vocabulary used (using English or Italian words, for instance) while sharing the same conceptualization”*. Notice that when defining ‘ontology’ in the CS reading, Guarino points to *a certain reality as the object of description*. In the philosophical reading, instead, the particular system of categories accounts *for a certain vision of the world*. Albeit it seems not he intended to bring in such a difference in these passages, it is in fact a not so subtle issue as old in Philosophy as Plato and Aristotle. Indeed, concept- or realism-orientation - which impinges in one or another theory of universals (42) - is a matter of discussion and have concerned ontologists in the biomedical ontology literature. We comment this issue in Section 3.3. At this point, it is only worth to say that this issue does not impinge any effect to our adopted ontology definition, since we stand for a more (terminologically) neutral Guarino’s definition (12, p. 6), viz., *“an ontology [is] a set of logical axioms designed to account for the intended meaning of a vocabulary”*.

It is still worthwhile to mention, however, that to build an ontology is the very task of *representing a domain*. Therefore, the intended meaning of a vocabulary, or the theory’s *ontological commitment* - first coined by Quine (37), should mirror it and nothing else but it.

### 2.1.3 How to Talk about Good Ontologies?

Guarino’s definition given above shed some light to the business of Ontology in Computer Science. A direct consequence of it, moreover, is that an ontology of a given non-trivial domain should constitute a highly-

axiomatized logical theory. This seems to be unavoidable whether it is to *constrain* the intended meaning of a vocabulary by harnessing the tool of logics. Perhaps it becomes more clear if we refer to Figure 4.

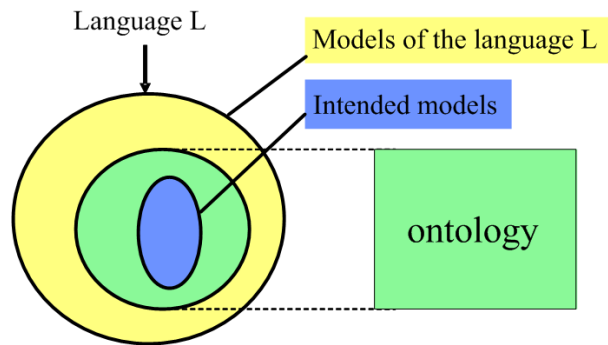


Figure 4: The intended models of a logical language reflect its commitment to a conceptualization. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating this set of intended models. Loosely adapted from (12, p. 7).

In Figure 4, one can figure that an ontology is as good as it approximates better the intended models inherent to the subject domain. The yellow area then indicates what you could say (represent) by means of the language L, i.e., the (possible) models of L. The blue area in turn delimits the subject domain itself, which is supposed to be covered in the models of L; notice that for this reason the language L must be expressive enough to afford that domain representation. Finally, the green area marks what we actually say in an ontology meant to represent the domain at hand.

Altogether, since it is hardly manageable to afford the *ideal ontology* of a given domain, we can say an ontology is likely a simplified view of it. Nevertheless, firstly, *an ontology has to comply with all the situations of the domain of study*<sup>1</sup>, i.e., the green area must cover the blue one in its entirety. Secondly, *the way to produce good ontologies is to strive for their best approximation to their correspondent subject domains*, i.e., the green area should get as close as possible to the blue one. As we have seen, it is hard to figure out a way to do that, if not by recurring to both (i) methodological principles to be employed along the ontology development, and (ii) a strong axiomatization capable of restricting the interpretation of what is said by the ontology to the domain itself (i.e., fitting the green area to the blue one).

## 2.2 Ontological Foundations

### 2.2.1 Formal Ontological Principles

In 2000, Guarino and Welty reported on a series of papers - e.g. (43) - how formal ontological principles such as *identity*, *unity*, *essence*, *dependence* and so forth can be used to support ontological decisions. Their aim has been to shift ontological practice in CS from an art to a rigorous engineering discipline founded on philosophical principles (44, p. 61). Indeed, the actual contribution ontological practice can provide in CS is to foster these principles into the common practice of knowledge engineering, as well as into conceptual modeling and domain modeling. Otherwise, the term ‘ontology’ in CS is bound to be “*simply a new word for something computer scientists have been doing for 20 - 30 years*” (44, p. 61).

<sup>1</sup>As where a domain exactly starts and ends can be a quite subjective issue, there is a need for setting objective lines (e.g., competence questions) under which a domain can be defined. Confer Section 2.4.



The work Guarino and Welty developed came to be called *OntoClean* (45), a methodology to “clean up ontologies”. In other words, *OntoClean* is meant to identify and correct flaws in the structure of ontologies specially by evaluating the *backbone taxonomy* formed by subsumption relationships. With the very single purpose of giving an illustration, consider the example below borrowed from (44, p. 62).

[...Consider] a proposed class time duration whose instances are things like “one hour” and “two hours”, and a class time interval referring to specific intervals of time, such as “1:00 - 2:00 next Tuesday” and “2:00 - 3:00 next Wednesday”. One proposal was to make time interval a kind of (subclass of) time duration, since all time intervals were seen as time durations. Seems to make intuitive sense, but how can we evaluate this decision? In this case, an analysis based on the notion of identity can be very informative. According to the identity criteria for time durations, two durations of the same length are the same duration. In other words, all one hour time durations are identical - they are the same duration and therefore there is only one “one hour” time duration. On the other hand, according to the identity criteria for time intervals, two intervals occurring at the same time are the same, but two intervals occurring at different times, even if they are the same length, are different. Therefore, the two example intervals given would be different intervals, but the same duration. This creates a contradiction: if all instances of time interval are also instances of time duration (as implied by the subclass relationship), how can they be two instances under one class and a single instance under another?

In fact, they can not; and this example turns out how the formal ontological principle of identity can support some ontological decisions. Many other examples involving other principles (e.g. unity, essence, etc) are given by Guarino and Welty in that notorious article to demonstrate the importance of an ontology engineering principled on philosophical foundations. In sum, they point out that structuring decisions should not result from heuristic considerations but instead should be motivated and explained in the basis of suitable ontological distinctions. Finally, their proposed methodology suggest the modeler to assign to the domain entities meta-properties that characterize their ontological behavior. The *OntoClean* methodology has been a pioneer initiative regarding the application of formal ontological principles in ontology development in CS. It has been used in conceptual modeling projects and integration efforts by several companies such as *OntoWorks* and *Document Development Corporation*, among others (46). For the sake of brevity, we have referred only to *OntoClean* in this section as it is one of the notorious efforts regarding the employment of formal ontological principles in ontology engineering. For a in-depth account of how such formal ontological principles in general, and *OntoClean* in particular, can be applied we refer the reader to (46).

## 2.2.2 Top-Level Ontologies

One of the reasons why Guarino and Welty’s work has been very influential is that it makes very little, if any, commitment to a particular ontology. It rather is based on techniques philosophers use to analyze, support, and criticize each others arguments. In point of fact, these techniques work very well for exposing what are often very subtle distinctions. In order to provide a framework of such a meta-ontological support, a number of *top-level ontologies* have been proposed in the ontology literature. Among them, we may cite Sowa’s ontology (47), Descriptive Ontology for Linguistic and Cognitive Engineering - DOLCE (48, 49), General Formal Ontology - GFO (50, 51), Basic Formal Ontology - BFO (52) and Unified Foundational Ontology - UFO (53). These top-level (or foundational, upper-level) ontologies share some ontological assumptions (e.g., the fundamental distinction between objects and processes), but disagree in others as well.

The fundamental ontological commitments and distinctions that are laid out in coherent top-level ontologies are part of the reason they can be useful in decision making during domain ontology development. Based on, say, the basic distinction between objects and processes, a number of axioms can be formulated that constrain what can

be stated in a specific domain about the interactions between its continuants (objects, or endurants) and occurrents (processes, or perdurants). For example, even though continuants can *participate in* occurrents (e.g., you are a participant in your life), continuants can not be *part of* occurrents (e.g., you are not part of your life) (26, p. 255). In summary, the adoption of a given top-level ontology to ground the development of a domain ontology does not actually push to the latter much of the philosophical assumptions of the former; but rather, impinges coherence between the domain modeling choices themselves. In the example just mentioned, the adoption of (say) BFO would not force the modeler to state that “you” exemplifies a continuant and “your life” an occurrent, but only that if it is said so, he/she should be aware that “you” can not be part of “your life”.

In this way, a top-level ontology can be used at the domain level to help with form, while keeping itself absent with respect to content<sup>2</sup>. Indeed, a top-level ontological framework comes not only to provide us with a support in making these decisions, but also to let these decisions as transparent as possible in the resulting domain ontology. This is because, if grounded in a top-level ontology, a domain ontology can be, say, annotated such that the meta-properties of the domain entities are explicitly marked on for the reader. In other words, a top-level ontological framework can also help in clarifying the domain ontology intended meaning.

A top-level ontology deals with the representation of such meta-properties and their relations. It thus consists a resource which a domain ontology can be grounded in. Thus, when committing to a particular top-level ontology, a domain ontology adheres to the domain-independent theories defined in the former. For example, if using (say) UFO (53), a statement saying that a Kind universal such as **Person** is subsumed by a Role universal such as **Student** turns to be inconsistent. The rationale behind this is that in Formal Ontology it is understood that a kind universal can not be subsumed by a role universal (46, p. 57). In the given example, for one, it is definitely not the case that every **Person** is necessarily a **Student**.

Altogether, Degen et al. state in (54, p. 34) that “*every domain-specific ontology must use as framework some upper-level ontology*”. Their claim for an upper-level ontology underlying a domain-specific one reflects the need for fundamental ontological structures, say, theory of parts, theory of wholes, types and instantiation, identity, dependence, unity, etc, in order to represent the domain properly. Similarly, in this thesis we draw attention to the fact that building a (biomedical) domain ontology on the basis of some top-level ontological framework is beneficial, if not necessary. In Subsection 4.1.1 we introduce UFO as the top-level ontology adopted along the development of the ECG Ontology.

## 2.3 Ontology Formalisms

The previous sections dealing with ontologies and their foundations have stressed the need for fostering highly-axiomatized ontologies as somewhat resembling a quality mark. Therefore, we have also seen that the language used to specify an ontology must be able to express the intended meaning of the vocabulary in hand. Such a language must then be expressive enough for allowing one to model the elements in the universe of discourse in terms of language constructs (or primitives). As examples of languages that qualify as suitable ontology formalisms for this purpose, we can cite First-Order Logic (FOL)<sup>3</sup>, or even more expressive logical formalisms such as modal or higher-order logics.

Nonetheless, one of the foremost motivations for ontology development is making use of automated reasoning.

<sup>2</sup>For this reason the term ‘Ontology’ (standing for the discipline) is often preceded by ‘Formal’, just as ‘Formal Logics’

<sup>3</sup>We are referring here to Predicate Calculus. For the rest of this text, we use ‘FOL’ with this meaning.

This is in fact one of the practical objectives of the work underlying this thesis. However, in general, the more expressive is a logical representation, the less efficient is an algorithm to process and retrieve information from it. Such a tradeoff has long been recognized by the knowledge representation community in AI - cf. (55) - and is still open as a topic of research. For this reason, common knowledge representation and deductive database languages - e.g., some instances of Description Logics (56) - have been specifically designed to afford decidability and efficient automated reasoning.

More recently, as part of the semantic web effort, many off-the-shelf ontology tools have been produced to support primarily query-based web applications supposed to access online ontologies. As a consequence of some good preliminary results, semantic web languages such as the Web Ontology Language (OWL) (57) have become popularized in such a way that the term ‘ontology’ has sometimes been mixed-up with the OWL format. In an invited talk in WebMedia 2008<sup>4</sup>, Giancarlo Guizzardi discussed why and how OWL-dogmatism can lead to a contradiction to the original purpose of using ontology in CS. Indeed, the OWL format does have its role in ontology development, but using it has practical implications which can not be ignored - cf. (58); and is only justifiable as a design choice. For instance, the OWL sub-language named OWL DL has been designed to ensure computational properties (viz., decidability and tractability) based on a given Description Logics (DL) family. In doing so, however, it has paid the price of sacrificing the expressiveness required to produce good ontologies (cf. Subsection 2.1.3).

In (59), Ceusters et al. demonstrate why to use only Description Logics (DL) for ontology development is not enough. They provide several examples of incorrect DL-based representations and then discuss how an ontology engineering principled upon philosophical foundations can prevent them to occur. However, while a disciplined use of the principles discussed on Subsection 2.2.1 covers yet the issues raised by Ceusters et al. (59), the point we are touching on here is something else. Namely, that beyond using formal ontological principles in terms of methodology along the development lifecycle, it is also beneficial to keep embedded in the ontology as much as possible of the assigned ontological properties and meta-properties of the domain entities. This is purposeful for (i) allowing the modeler to be explicit regarding his/her ontological commitments (33); this enables him/her to expose subtle distinctions between possible models, which is useful in ontology integration for minimizing the chances of running into a *False Agreement Problem* (12); (ii) support the modeler in justifying his/her modeling choices and providing a sound design rationale for choosing how the elements in the universe of discourse should be modeled in terms of language elements.

According to Guizzardi and Guarino (60), an ontology formalism ought to support modelers in creating specifications which are as truthful as possible to the domain being represented (*domain appropriateness*) and as efficient as possible in supporting communication, understanding / learning and problem solving about that domain (*comprehensibility appropriateness*). With this purpose, to emphasize it once more, the ontology formalism used must to be expressive enough. As an attempt to make this tangible, consider the example shown in Figure 5. It contrasts the quality of (i) two models designed with no philosophical foundation in standard UML with (ii) a model designed with the support of the top-level ontology UFO in an ontologically well-founded UML profile. It may be worth mentioning that the former models are found often in the specification of information systems (58).

Indeed, in case one uses an ontology formalism that has embedded ontological distinctions in its language constructs (as the stereotypes do in Figure 3.c), models that represent 3.a and 3.b assertions become syntactically

<sup>4</sup>See the abstract of Guizzardi’s presentation entitled “OWL-dogmatism considered harmful: The role of foundational ontologies for the Semantic Web” ( [http://www.inf.ufes.br/webmedia2008/webmedia2008\\_keynote.html#Giancarlo](http://www.inf.ufes.br/webmedia2008/webmedia2008_keynote.html#Giancarlo)) in the 14th Brazilian Symposium on Multimedia and the Web.

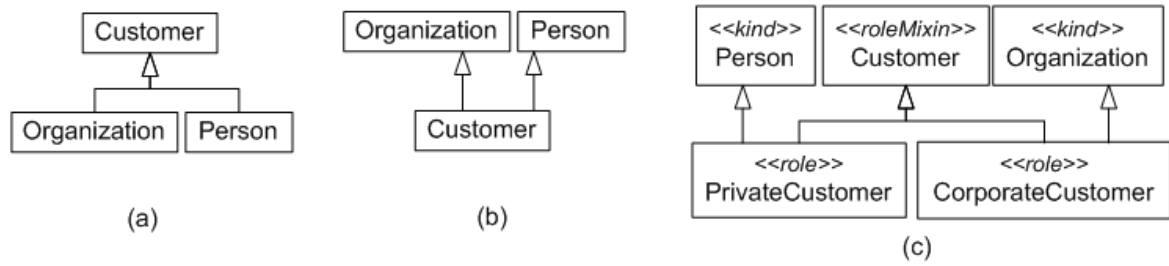


Figure 5: Three different possible models for representing the concept of customer, which can be either a person or an organization. The models 3.a and 3.b are ontologically incorrect since: (i) in 3.a, it is not the case that all instances of person (or organization) are customers; (ii) according to 3.b, every instance of **Customer** is both **Person** and **Organization**, thus, the extension of **Customer** is empty. The model 3.c, otherwise, is a design pattern that provides an ontological solution to this recurrent problem in CM (58, p. 23). In UFO, while an instance of Kind universal (e.g., **Person**) must be always distinguished as such by holding a well-defined identity, an instance of Mixin universal (e.g., **Customer**) must not, since its instances can be of different kind universals (e.g., **Person**, or **Corporation**). Hence, both **Private Customer** and **Corporate Customer** (which are Role universals) can be said (disjoint) types of **Customer**, a Mixin universal. UFO is introduced with more detail in Subsection 4.1.1.

incorrect. Therefore, in this case not only ontological foundations have been used to support modeling decisions, but the ontological distinctions between concepts have been made explicit to the reader of the model.

On the other hand, a higher-order statement such as one saying that a given individual is an *instance of Private Customer* which is an *instance of Role*, in general, jeopardizes the computational properties of decidability and tractability of a deductive system. With this in mind, Guarino underlies in (12, p. 9) that we actually need two kinds of ontologies, viz., coarse and fine-grained ontologies. On the one hand, the latter is meant to actually “be” the ontology of a given domain, i.e., that one which “*gets closer to specifying the intended meaning of a vocabulary (and therefore may be used to establish consensus about sharing that vocabulary)*”. On the other hand, the former is intended to be “*...a minimal set of axioms written in a language of minimal expressivity, to support only a limited set of specific services, intended to be shared among users which already agree on the underlying conceptualization*”. Guarino has called the first kind mentioned *reference ontologies*, or off-line ontologies; and the second one *shareable ontologies*, or online ontologies. In this thesis we shall refer to the former in the same way, whereas to the latter as *lightweight ontologies*.

Thomas Bittner and Maureen Donnelly have in turn drawn attention to an analogous call for, viz., that two different kinds of *formalisms* are required for ontology development. They first advocate that highly-expressive languages (they choose FOL) are required to properly represent ontological theories. They say that such representations should be carried out “*in a single deductive system that is expressive enough to make critical distinctions in logical properties explicit*”. They, however, recognize as well the need for using DL to further convey *lightweight* versions of the original ontology by selecting a specific DL formalism for it. The motivation for this is to exploit DL only in what they do can offer, i.e., be very valuable and capable tools for computational ontologies that support effective automated reasoning. In parallel, Guizzardi has elaborated, first in (33), and also in (61, p. 8), on a systematic ontology engineering approach that takes this need for two ontology formalisms into consideration. In the next section we discuss this approach, which has been employed in our research.

## 2.4 Ontology Engineering

In addition to all the foundational issues that make Ontology a purposeful discipline in Computer Science, methodological guidelines in terms of engineering (i.e., ontology development lifecycle, use of supporting tools and so forth) have also been a topic of research. Uschold and Gruninger seem to be the first to propose guidelines for ontology building, by relying on their experience in developing the Enterprise Ontology (62). They point to some key processes to be carried out, viz., (i) identify the ontology's purpose, (ii) build the ontology, (iii) evaluate the ontology, and (iv) document the ontology. They highlight ontology capturing as the main task in ontology building, which concerns the identification and definition of key concepts and relationships in the universe of discourse. These guidelines have been the main source for the ontology building method named SABIO, proposed by Ricardo Falbo in (63). The effectiveness of SABIO has been tested for more than ten years in the development of a number of domain ontologies in areas ranging from Harbor Management to Software Process to Media on Demand Management.

Perhaps the foremost point of innovation Uschold and Gruninger have brought in is the introduction of the so-called *competence questions* (CQ). These questions are meant to be used as a means for both identification of the ontology's purpose and scope and a testbed for evaluation. The CQ's are intended to be formulated in formal logic for delimiting an objective criterion for discussing the ontology effectiveness and completeness as long as the CQ's are demonstrated to be answered.

Another point of discussion in Ontology Engineering is whether or not an ontology can represent a domain, independent of specific application concerns. On one side, Nicola Guarino advocates in (64) that a domain ontology - not an *application ontology* (64, p. 300) - can and should represent domain knowledge<sup>5</sup>. He sustains that only the level of granularity of the domain knowledge used to build an ontology is dependent on the particular concerns the ontology in hand is made for. Under this vision, Guarino claims that ontology development favors a systematic quest for reusability. On the other side, Mustafa Jarrar (65) echoes van Heijst et al. (66) in stating that any ontology is biased by specific concerns that motivate the ontology construction (the so-called *interaction problem*). Jarrar then argues that there is a tradeoff between ontology reusability and usability, and point out that a balance should be pursued between these two conflicting issues.

As we have discussed in the introduction of this thesis, we shall provide evidence here that an ontology representing a domain (in the terms presented in Subsection 2.1.3 above) can be used to derive a lightweight ontology (it can be interpreted as "application ontology") with genuine benefits, cf. Goal 3 and Research Question 3. In other words, we expect to demonstrate throughout this thesis that a reference domain ontology can be said at the same time *reusable* and *applicable* (or usable), even if the latter requires some adaptation to computing issues.

Along these lines, to bring back the point mentioned in the end of the previous section, we reflect in this thesis a systematic ontology engineering process proposed by Guizzardi (33, 61) that comprises the phases of analysis, design and implementation in ontology construction. Basically, in the first phase a reference ontology is to be created, independent of codification language and application concerns. This model is intended to be as truthful as possible to the domain. Subsequently, with the aim of addressing a specific application, a design phase is required for choosing a codification language that meets the application non-functional requirements. The same reference ontology can then give rise to different ontology codifications in different languages (e.g., F-Logic, OWL DL, RDF, ORM, Ontolingua) laid in the solution space. Finally, in the implementation phase, the reference

<sup>5</sup>A discussion on the (possible) distinction between a *domain* and *domain knowledge* is given in Section 3.3.

ontology is specified in the chosen codification language. This ontology implementation (the so-called lightweight ontology) is intended to be an online model, amenable to be used (say) in knowledge-based systems for inference and information retrieval purposes.

## 2.5 Conclusions

To conclude this chapter, we report to the following quotation from Alexander Yu (26, p. 264), to echo that

Philosophical ontology has much to offer in terms of formal analytical methods towards creating declarative representations of knowledge that are general, reusable, and valid. At the same time, we need to also draw upon the insights and approaches that have developed within the engineering community, particularly those that have exposed and attempted to address practical problems that continue to dog both users and developers of ontologies.

In this spirit, the key points we draw attention to are:

- Although still up to this point there is controversy regarding ontology definition and methodologies in CS, pioneer researches have been working heavily on making Ontology a meaningful discipline in CS. As a result, key quality principles for ontologies have been already established.
- Ontological foundations are very valuable, if not indispensable tools to set the ground for ontology development. They provide the modeler with a support to make critical ontological decisions in the representation of a given domain. Moreover, such ontological decisions can be kept embedded in the models produced to increase the level of transparency concerning the assumptions made. This, however, requires the ontology formalism used to be expressive enough to make the necessary distinctions.
- In face of the tradeoff between expressiveness and computational tractability, to cite one primary reason, ontology development is also about making use of engineering tools. This is because, in order to meet conflicting requirements of different artifacts to be produced in different phases of an engineering process, specific properties should be focused on in deference to others.

We now are able to move to *biomedical ontology*, a research field that has been emerged in the context of Medical and Bioinformatics.

## 3 Background, Part II: Biomedical Ontology

This chapter provides some background on the biomedical ontology literature. By surveying it, one comes across several open issues of methodological, philosophical and computational nature. Among the most salient, we might cite (i) the pursuit of conveying well-defined biomedical entities (67, 16); (ii) an ongoing controversy about concept- vs. realism-orientation (68, 27, 28, 69); (iii) the challenge of making the definition of parthood relations in Biomedicine as mature as subsumption (70); (iv) the use of core biomedical ontologies to serve as top-domain frameworks for supporting ontologies dealing with specific sub-domains in Biomedicine and also help in their integration (71, 72); and (v) the need for handling default knowledge to afford integration of canonical ontologies (that consider an idealized view on a domain) and phenotype ontologies (that take account of properties or phenomena, when exemplified by individuals) (9, 73).

A complete consideration of all these topics is beyond the scope of this thesis. We, however, give an account of some of them which we deem relevant in this text. The chapter then starts with an overview of referred biomedical terminologies and ontologies. In the sequel, we comment on how these developing artifacts fit in the formal ontological principles discussed in the previous chapter. We then provide a brief account of the controversy of concept- vs. realism-orientation as far as it has been influential in the conduction of our research. After that, we give an overview of how biomedical ontologies have been applied up to this point. Finally, we outline some conclusions and a recapitulation of the key points developed in the course of this chapter.

### 3.1 Biomedical Terminologies and Ontologies

Physicians have developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. In the end, such clinical vocabularies, terminologies or coding systems are intended to convey *terms* for describing unambiguously the care and treatment of patients. Terms cover diseases, diagnoses, findings, operations, treatments, drugs, administrative items etc., and can be used to support recording and reporting a patient's care at varying levels of detail, whether on paper or, increasingly, via an electronic health record (EHR).

Before we introduce some of the main existing biomedical terminologies, there are some important notions that must be clarified. Consider the definition given by ISO for *terminology* as put by Schulz et al. in (74).

A terminology is a set of terms representing the system of concepts of a particular subject field. Terminologies relate the senses or meanings of linguistic entities with concepts. Concepts are conceived as the common meaning of (quasi-)synonymous terms (75).

In other words, terminologies focus on terms, which are their unities of information. In a terminology, the

purpose of a definition is then to outline all meanings associated with a given term. For instance, according to the Merriam-Webster's online dictionary, the term 'head' may refer, among other things, to "*the upper or anterior division of the animal body that contains the brain, the chief sense organs, and the mouth*", or to "*one in charge of a division or department in an office or institution*".

By taking our definition of 'ontology' given in Section 2.1 into account, in an ontology, instead, the unit of information is an entity in reality. Hence, in an ontology the purpose of definitions is to *precisely delimit the possible interpretations of these entities*. Such a difference has in fact practical implications that have been outlined in a number of articles, e.g. (28, 76, 74). Specially in (28), it is demonstrated the importance of focusing not on the representation of terms, but on entities in reality. Nonetheless, in Medical Informatics many language-centered concept systems have been developed, albeit their widespread adoption has been slow. Among the most referenced biomedical terminologies in the literature, we can cite UMLS Semantic Network (1), NCI Thesaurus (2) and SNOMED-CT (77). They have been designed to meet different and specific goals, varying in their coverage and completeness. As follows we provide a brief account and criticism of these terminologies as it can be found in the biomedical ontology literature. After that, we turn to the Gene Ontology and the Foundational Model of Anatomy which the literature refer to more often as two biomedical ontologies.

### 3.1.1 UMLS Semantic Network

The Unified Medical Language System (UMLS)<sup>1</sup> aims "*to facilitate the development of computer systems that behave as if they 'understand' the meaning of the language of biomedicine and health*" (1). It is composed by three different parts (Ibid.):

- *The Meta-thesaurus*, which contains over one million biomedical concepts from over 100 source vocabularies;
- *The Semantic Network*, which defines 135 broad categories and fifty-four relationships between categories for labeling the biomedical domain;
- *The SPECIALIST Lexicon & Lexical Tools*, which provide lexical information and programs for language processing.

Our interest here is only on the UMLS semantic network, which has been evaluated as a biomedical ontology (26, p. 258). The semantic network provides a categorization of the concepts (called *semantic types*) present in the UMLS Meta-thesaurus and a set of *relationships* between these concepts. The current release (November 2008) of the semantic network contains 135 semantic types and 54 relationships. The information associated with each semantic type includes (1): (i) a unique identifier, (ii) a tree number indicating its position in the 'is-a' hierarchy, (iii) a definition and (iv) its immediate parent and children. In turn, the information associated with each relationship includes the items (i), (ii) and (iii) above, and additionally (v) the semantic type, (vi) examples and (vii) the set of semantic types that can plausibly be linked by this relationship.

Examples of UMLS semantic types are organisms, anatomical structures, biologic function, chemicals, events, physical objects, and even concepts or ideas. The UMLS taxonomy is organized in two main categories, viz., *Entity* (e.g., amphibian, gene or genome, carbohydrate) and *Event* (e.g., social behavior, laboratory procedure, mental process). Figure 6 gives an illustration of a portion of the UMLS semantic network.

<sup>1</sup>From US National Library of Medicine, available at: <<http://www.nlm.nih.gov/research/umls/>>.



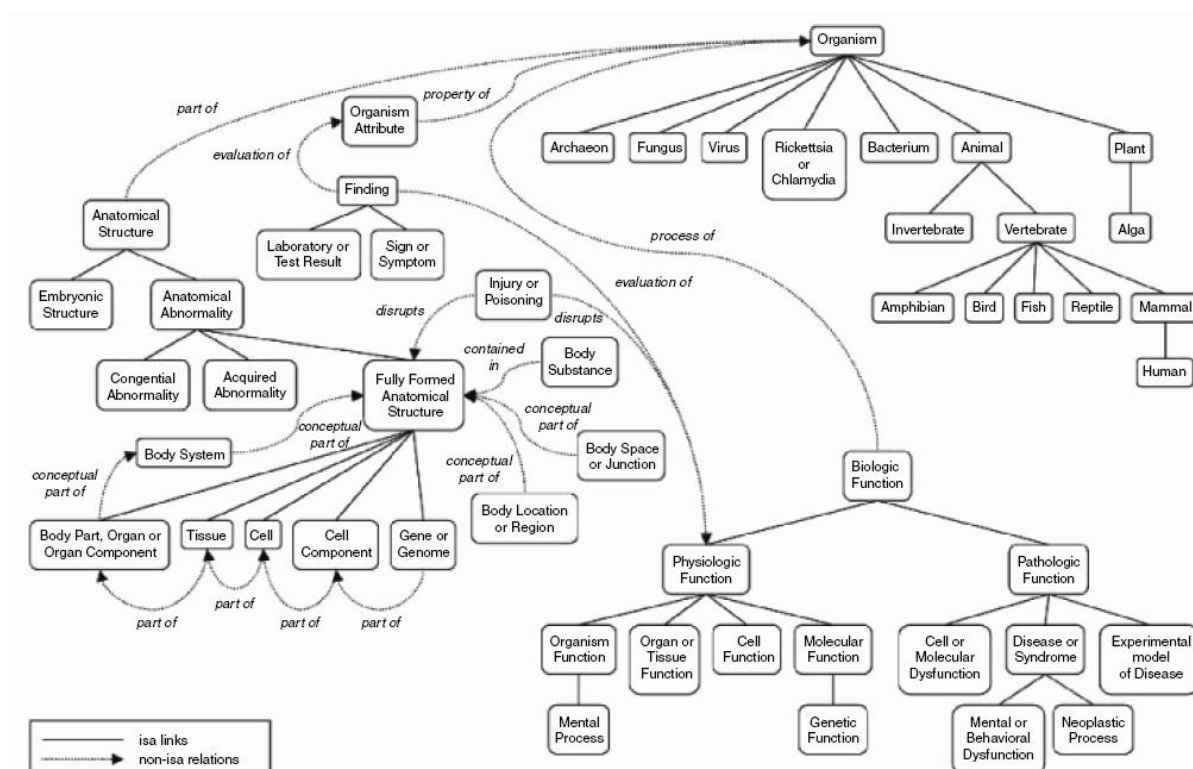


Figure 6: Portion of the UMLS semantic network - source: (1).

In (78), McCray introduces the UMLS semantic network as a biomedical ontology. As such, it has been analyzed with respect to its ontological soundness. As a result of an evaluation conducted by Sculze-Kremer et al. (79), many revisions have been suggested to correct structural problems. The authors point, for instance, to the UMLS statements *plant roots is-a plant* and *plant leaves is-a plant* to demonstrate the UMLS misleading mixing-up *is-a* and *part-of* relations, since it is not the case that the *is-a* relation holds above, but rather *part-of*. Also in (80), Kumar and Smith provide a critical account of the UMLS semantic network in the light of the formal ontology BFO (52). One example given by them to demonstrate inconsistencies caused by the UMLS ambiguity refers to the term ‘cardiac output’. It (exotically) is, according to the UMLS semantic network, both a continuant (or enduring) and occurrent (or perdurant) - ontological categories which are fundamentally disjoint. In face of these criticisms, it turns out that the UMLS semantic network needs a major review if it is to be regarded as a biomedical ontology.

### 3.1.2 NCI Thesaurus

The US National Cancer Institute’s (NCI) thesaurus<sup>2</sup> is a public domain controlled vocabulary created by the cancer research community. It evolved from the NCI Meta-thesaurus, which is in turn based on the UMLS Meta-thesaurus. The NCI thesaurus is an effort to integrate molecular and clinical cancer-related information, mostly on the side of terminologies. It is DL-based and comprises definitions for basic and clinical concepts used in cancer research, a taxonomic structure of these concepts and relations between concepts as well (81, 82). Figure 7 depicts an exemplary query result given by the NCI thesaurus for the term ‘tumor-derived’.

The authors of the NCI thesaurus claim “it is deep and complex compared to most broad clinical vocabularies,

<sup>2</sup><http://nciterns.nci.nih.gov/NCIBrowser/Dictionary.do>.

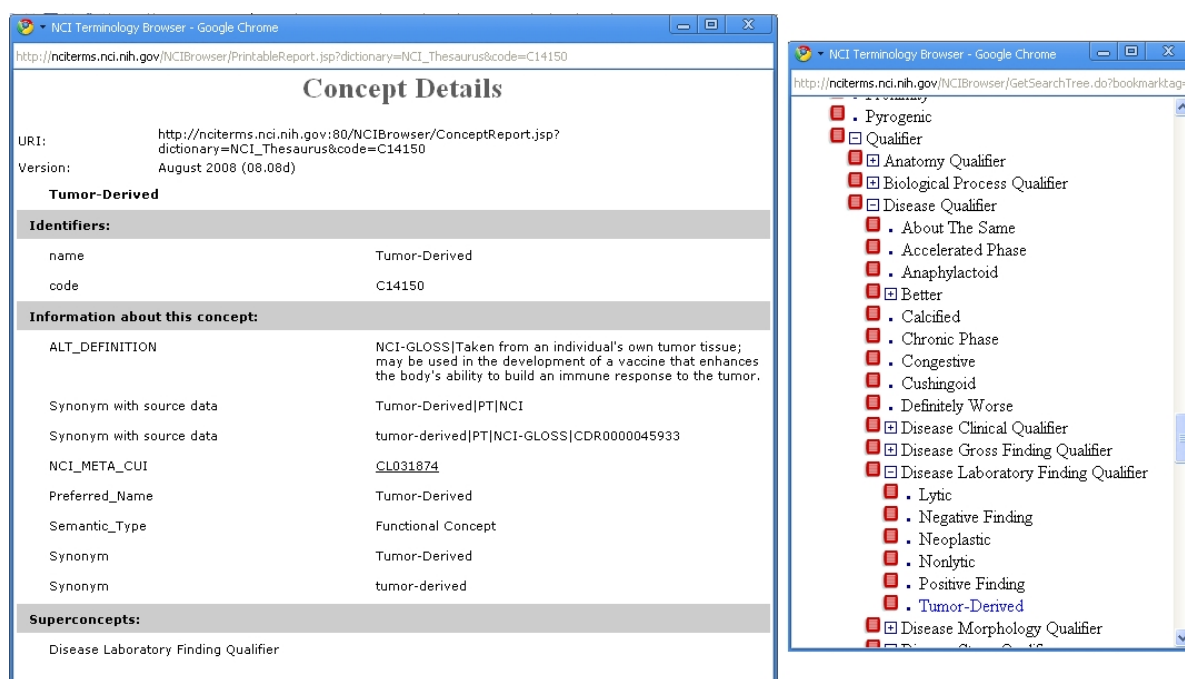


Figure 7: A NCI thesaurus' query result on the term 'tumor-derived' (2).

implementing rich semantic interrelationships between the nodes of its taxonomies.” (81). It, however, has been pointed as suffering from the same broad range of problems that have been observed in other biomedical terminologies. It has been analyzed by Ceusters et al. (83) in the light of both ISO terminology standards and ontological principles advanced in the recent biomedical literature. As a result, many problems were found, ranging from mistakes and inconsistencies regarding the term-formation principles used to missing or inappropriately assigned verbal and formal definitions. Moreover, other analysis has been made by Kumar and Smith (84) as an attempt to assess the integration of data and information deriving from different sub-domains of Oncology covered by the NCI thesaurus. They have focused on the kinds of entities which are fundamental to an ontology of colon carcinoma. Likewise, they report problems found with respect to classification, synonymy, relations and definitions<sup>3</sup>. Nonetheless, Kumar and Smith state that “*The NCI thesaurus does provide a rich terminology for carcinomas, which makes it a good starting point for ontology work in the cancer domain*” (84). They then propose means for the repairs needed to qualify the NCI thesaurus as a “reference” ontology for the cancer domain in the future.

### 3.1.3 SNOMED-CT

SNOMED-CT<sup>4</sup> (Systematized Nomenclature of Medicine - Clinical Terms) is a multilingual clinical healthcare terminology. The motivation of the SNOMED-CT curators is that “*the delivery of a standard clinical language for use across the world's health information systems can [...] be a significant step towards improving the quality and safety of healthcare*” (77). They believe such an initiative could avoid deaths and injuries that are results of poor communication between healthcare practitioners. They also consider that, ultimately, patients would benefit from the use of SNOMED CT, since it builds and eases communication and interoperability in electronic health

<sup>3</sup>Kumar and Smith suggest that, in adhering to its legacy in the UMLS Semantic Network, the NCI thesaurus has increased the number of its inaccuracies (84, p. 219).

<sup>4</sup><http://www.ihtsdo.org/snomed-ct/>.

data exchange (77).

A pertinent quotation from (74) provides us the current tendencies of SNOMED-CT:

- the urgent need for a global standardized terminology for medicine and life sciences, suitable to cope with an immense flood of clinical and scientific information;
- an impressive legacy of systematized biomedical terminology;
- efforts toward an ontological foundation of the basic kinds of entities in the biomedical domain as an important endeavor of the emerging discipline of “Applied Ontology” (85);
- the increasing availability of logic-based reasoning artifacts suited for large ontologies.

SNOMED-CT covers the core general terminology for the electronic health record (EHR). It currently contains over 311,000 concepts with unique meanings and description logic-based definitions hierarchically organized (77). The concepts are linked to terms and multi-lingual synonyms. Their meaning can be obtained both from their position in the hierarchy and from formal axioms that connect concepts across the hierarchies (74). When implemented in software applications, SNOMED-CT can be used to represent clinically relevant information as an integral part of producing electronic health records (77). Figure 8 depicts a SNOMED-CT’s tree view and a query result for the concept of “Record artifact”.

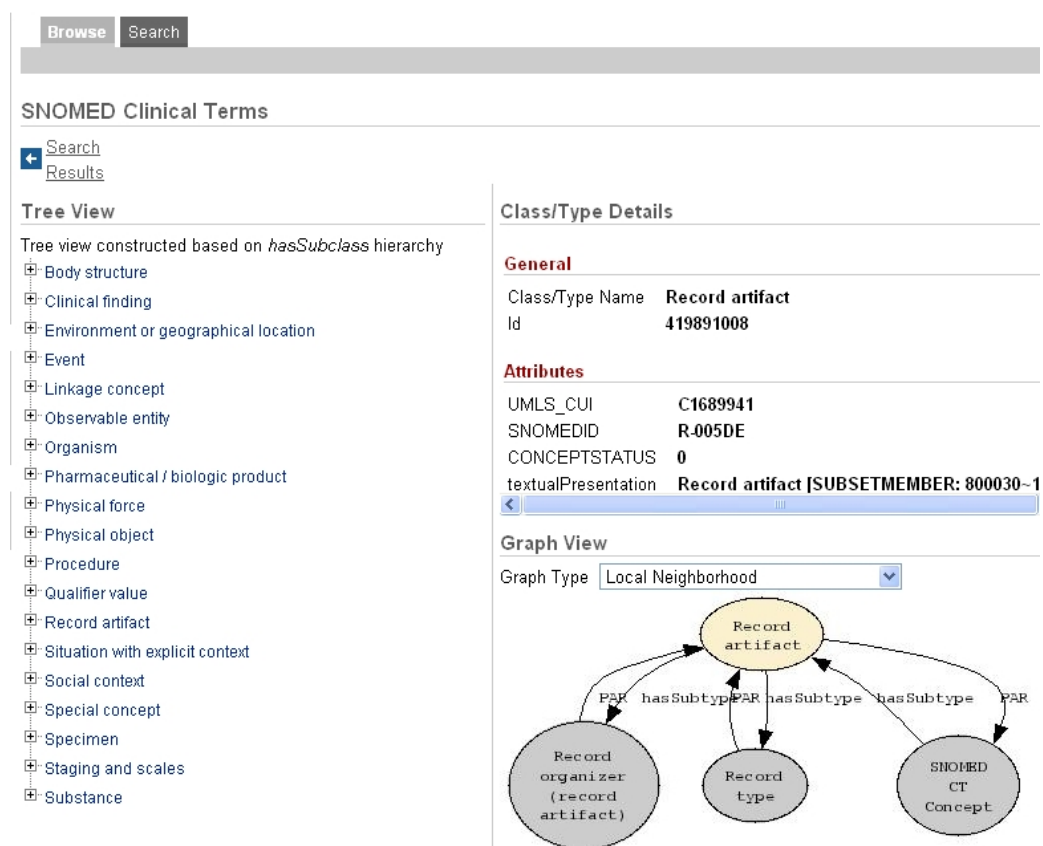


Figure 8: SNOMED-CT’s tree view and a query result for the concept of “Record artifact” (source: <<http://bioportal.nci.nih.gov/>>).

The SNOMED-CT history and usage over the last forty years has been analyzed by Cornet and Keizer in (86). Their conclusions are as follows.

The clinical application of SNOMED is broadening beyond pathology. The majority of studies concern proving the value of SNOMED in theory. Fewer studies are available on the usage of SNOMED in clinical practice. Literature gives no indication of the use of SNOMED for direct care purposes such as decision support.

Like most of the existing biomedical terminologies, SNOMED-CT has been reviewed in terms of its ontological soundness. In (59), Ceusters et al. demonstrate through an analysis of SNOMED-RT (an earlier version of SNOMED-CT) that description logics is not enough to foster sound biomedical terminologies (or ontologies). They discuss several terminological and ontological problems present in SNOMED-RT. The most serious and recurrent, however, is the subsumption misuse. As in the UMLS semantic network, the is-a relation is mixed-up with the mereological one, cf. (59). Subsumption misuse in SNOMED-CT is reported as well by Bodenreider et al. in (76). A more recent SNOMED-CT review can be found in (74), where Schulz et al. point out SNOMED-CT is now reaching its adolescence and then asks for an ontologist's and logician's health check (in a clinician's jargon). They then propose a number of "therapeutic principles" to be applied for SNOMED-CT's long-lasting fitness and its increasing ability to stand the upcoming challenges of medical documentation and standardization.

As a result of such criticisms, UMLS, the NCI thesaurus, SNOMED-CT and other biomedical terminologies are gradually evolving from relatively *"simple code-name-hierarchy structures, into rich[er], knowledge-based ontologies of medical concepts"* (87). Although all of these three artifacts are inherently language- or term-centered, the real reason why we have called them 'terminologies' is just to bring in what the current literature portends. Indeed, there seems to be no precise frontier to classify a biomedical domain representation either as a terminology or as an ontology (31, p. V). The transition from the former to the latter is rather a dense path that is traced as long as formal ontological principles are more and more followed. This being said, we proceed to consider the Gene Ontology as subject of our analysis.

### 3.1.4 Gene Ontology

The Gene Ontology Consortium is a joint project that began from gathering three model organism databases to produce a structured, precisely defined, common, controlled vocabulary for describing the roles of genes and gene products in any organism (18). Therefore, in spite of what the name suggests, the Gene Ontology (GO) can be hardly said an ontology, but in fact a "controlled vocabulary" (88).

The three initial GO databases, viz., the FlyBase, Mouse Genome Informatics (MGI) and *Saccharomyces* Genome Database (SGD), are still to be combined to other model organisms' sources (18). As of June 19, 2003, GO was containing 1297 component, 5396 function and 7290 process terms. The total number of GO informal term definitions was 11020. Those terms are organized in hierarchies indicating either that one term is more general than another or that the entity denoted by one term is part of the entity denoted by another (88).

Database compilers or human curators annotate terms standing for genes or gene products in their databases in order to describe the processes in which the latter are involved. As it can be noticed, GO's focus is not ontological in either sense of philosophical ontology nor the information / computer science one introduced in Section 2.1. In other words, neither logically rigorous formalization and representational adequacy that provide stability for an ontological framework and its extendibility in the future, nor reasoning efficiency even at the price of simplifications on the representational side has been paid attention in the GO Consortium. Instead, its very

purpose is to provide a practically useful framework for keeping track of the biological annotations that are applied to gene products (89).

In GO, terms are divided into three disjoint trees, or sub-ontologies: *Cellular Component*, *Biological Process* and *Molecular Function*. A gene product might be associated with or located in one or more cellular components; it is active in one or more biological processes, during which it performs one or more molecular functions (3, access on March 03). For example, the gene product ‘cytochrome c’ can be described by the molecular function term ‘oxidoreductase activity’, the biological process terms ‘oxidative phosphorylation’ and ‘induction of cell death’, and the cellular component terms ‘mitochondrial matrix’ and ‘mitochondrial inner membrane’.

The task of synthesizing what each one of those three sub-ontologies (at the same time, three terms) stand for is not easy. This is because, even in the publications of the GO Consortium, there is controversy in their textual definitions. Nevertheless, we put below a those definitions as they are provided in the main GO sources.

- *Cellular Component*: refers to the place in the cell where a gene product is active (18, p. 27). Cellular component includes terms like ‘ribosome’ or ‘proteasome’, specifying where multiple gene products would be found. It also includes terms such as ‘nuclear membrane’ or ‘Golgi apparatus’. This may be an anatomical structure (e.g. ‘rough endoplasmic reticulum’ or ‘nucleus’) or a gene product group (e.g. ‘ribosome’, ‘proteasome’ or a ‘protein dimer’) (Ibid.).
- *Biological Process*: a series of events accomplished by one or more ordered assemblies of molecular functions (3, access on March 03). Processes often involve a chemical or physical transformation, in the sense that something goes into a process and something different comes out of it (18, p. 27). Examples of broad (high level) biological process terms are ‘cellular physiological process’, ‘cell growth and maintenance’ or ‘signal transduction’. Examples of more specific (lower level) terms are ‘translation’, ‘pyrimidine metabolic process’ or ‘alpha-glucoside transport’ (Ibid.).
- *Molecular Function*: the biochemical activity (including specific binding to ligands or structures) of a gene product (18, p. 27). This definition also applies to the capability that a gene product (or gene product complex) carries as a potential. It describes only what is done without specifying where or when the event actually occurs (Ibid., p. 27). Molecular functions generally correspond to activities that can be performed by individual gene products, but some activities are performed by assembled complexes of gene products (3, access on March 03). Examples of broad functional terms are ‘enzyme’, ‘transporter’ or ‘ligand’. Examples of narrower functional terms are ‘adenylate cyclase’ or ‘Toll receptor ligand’.

Figure 9 provides a screenshot of the AmiGO system, where the GO database can be accessed. Figure 10 in turn depicts the term information for ‘biological process’.

Recently, there have been several papers that provide a critical analysis of GO. We report to Smith et al. (88) and to Kumar and Smith (80) with this purpose. Both of them discuss some of the main problems encountered in GO and propose solutions in the light of Ontology to improve GO’s organizing principles. Among the flaws identified, we can cite:

- GO curators state that the three sub-ontologies aforementioned are disjoint; instead, there are strong evidence for *part-of* relationships holding between elements of distinct trees (88, p. 110), (80, p. 147), namely, between the realization of some functions and broader biological processes in which they unfold.

Figure 9: The three sub-ontologies of the Gene Ontology (3).

## biological\_process

Figure 10: A GO's query result on the term 'biological process' (3).

- the molecular function sub-ontology requires a comprehensive review. There are terms such as 'anti-coagulant' (defined as: "a substance that retards or prevents coagulation") and 'enzyme' (defined as: "a substance... that catalyzes") which are conveyed as molecular functions, but are rather substances, not functions (88, p. 110), (80, p. 146). To partially overcome that problem, GO curators have appended

‘activity’ to almost all terms under ‘molecular function’; however, as the term definitions have not been changed accordingly, other inconsistencies took place.

- the *is-a* relation is used with no precise meaning in GO; the GO documentation states that *is-a* stands for instance-of in GO, though it is broadly used as the *subsumption* (also termed *kind-of*, or *class-of*) relationship. However, there are examples in GO where *is-a* is taking the place of *part-of*, or even examples where *is-a* stands as a non-necessary subsumption, as if it were not to mean that every instance of the subsumee is subsumed by an instance of the subsumer, cf. (88, p. 112).
- the *part-of* relation is intended in GO to denote “can be a part of”, rather than “is always a part of” (88, p. 112). Moreover, this general relation is used in different cases which parthood distinctions are required if GO is to provide a sound domain representation (Ibid., p. 113). Additionally, *part-of* is always set for transitivity regardless whether or not the specific type of *part-of* in hand is in fact transitive - the transitivity problem in Conceptual Modeling is revisited by Guizzardi in (90).

Overall, there are several issues of concern which deserve a strong effort on from the part of GO curators. In case a methodology such as OntoClean is applied to GO, its consistency and coherence and thus its future applicability in the automated processing of biological data might be enhanced (80). There is an ongoing project entitled Gene Ontology Next Generation (GONG) (91, 92) that, though bringing the benefits of migrating GO to a more rigorous status for affording OWL DL-based automated reasoning, is not concerned with any review of the GO’s ontological status.

Nevertheless, the increasing significance attained by GO can be evidenced by the fact that, as of December 2004, there have been close to 300 articles in PubMed referencing GO (93, p. 2). In that article, Suzanne Lewis analyzes that “GO has succeeded because it is not a technical solution *per se*”. She adds that

“We want to continue integrating our knowledge forever and technologies are short-lived. So, the solution must be to adopt new technologies as they arise while the primary focus remains on cooperative development of semantic standards: it’s about the content, not the container.”

Nonetheless, we believe it is worth mentioning that, although a remarkable advance in genomic data integration that opens promising perspectives, GO still suffers from serious ontological problems to be overcome in its *content*, which by the way is also in virtue of its *container*.

### 3.1.5 Foundational Model of Anatomy

Initially developed as an enhancement of the anatomical content of UMLS, the Foundational Model of Anatomy (FMA) (94, 7) deals with the structure of the mammalian (specially the human body). It comprises the material objects from the molecular to the macroscopic levels that constitute the body and associates with them non-material entities (spaces, surfaces, lines, and points) required for describing structural relationships. This anatomy ontology is intended as a reusable and generalizable resource of deep anatomical knowledge, which can be filtered to meet the needs of any knowledge-based application that requires structural information (7).

The FMA has been developed throughout ten years in the School of Medicine of the University of Washington. It currently contains approximately 75,000 distinct anatomical concepts representing structures ranging in size from some macromolecular complexes and cell components to major body parts. These concepts are associated with over 120,000 terms. For each concept, multiple synonyms are assigned. The concepts are related to

one another by over 2.1 million relationship instances of over 168 relationship kinds (94). The FMA is an evolving ontology, which is one of the largest computer-based knowledge sources in the biomedical sciences. It is implemented in a frame-based system and is stored in a relational database (7). Figure 11 presents a screen shot of the Foundational Model Explorer (FME) system used to retrieve information efficiently from the FMA at (94).

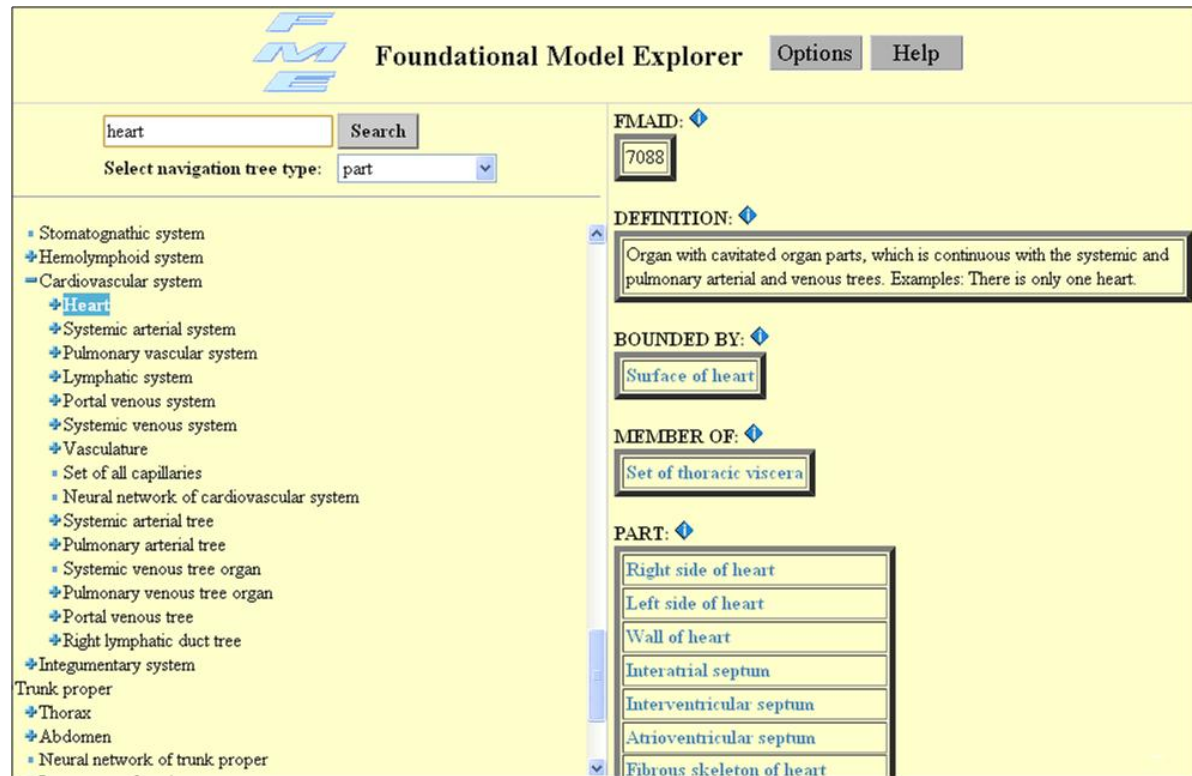


Figure 11: The Foundational Model of Anatomy illustrated by the FM Explorer available at <http://bioportal.nci.nih.gov/>. Notice on the left side of the figure, a part of FMA's partonomy with the concept of **Heart** selected. On the right-hand, it is shown the concept unique ID, in addition to its textual definition and the relationships held with other concepts which are laid in frame-based slots.

The FMA has four interrelated components, namely (94):

- Anatomy taxonomy (At): classifies anatomical entities according to the characteristics they share (genus) and by which they can be distinguished from one another (differentia) - designated in previous publications as the Anatomy ontology or Ao;
- Anatomical Structural Abstraction (ASA): specifies the part-whole and spatial relationships that exist between the entities represented in At;
- Anatomical Transformation Abstraction (ATA): specifies the morphological transformation of the entities represented in At during prenatal development and the postnatal life cycle;
- Meta-knowledge (Mk): specifies the principles, rules and definitions according to which classes and relationships in the other three components of FMA are represented.

The FMA is thus referred to by means of the abstraction  $FMA = (At, ASA, ATA, Mk)$ .



The Foundational Model of Anatomy covers a domain which is very basic in Biomedicine, and this is one of the reasons why it has been likely the most applied biomedical ontology in the literature. Indeed, anatomy is a well-studied domain in Biomedicine. If on the one hand, this makes easier the task of looking for knowledge sources and empirical evidences to rely on, on the other hand it requires a strong effort to produce an ontology that does justice with that domain in its entirety. In this respect, the FMA has been a result of a noteworthy work.

Nonetheless, particularly the challenge of coping with multiple levels of granularity that coexist is tough, and as pointed out by Kumar et al. (89, p. 505) and Rector et al. (95, p. 345), the FMA needs a thorough restructuring to address this issue satisfactorily. In addition, the assignment of *part-whole* relations in FMA has also been a concern for ontologists. In (96), Donnelly et al. provide an ontological analysis of the use of this relation in FMA. They demonstrate that the FMA collapses at least three different distinctions of parthood which are in general relevant in anatomy. This problem has far-reaching consequences in transitivity reasoning over FMA's *partonomy* (parthood taxonomy)<sup>5</sup>, since transitivity is allowed in every part-of relation even if it does not hold. However, the FMA is represented (only) in a frame-based framework that, although fitting as an ontology codification framework in virtue of a design decision, falls short in serving as a reference ontology specification framework due to its low expressiveness. An immediate consequence is, for one, that FMA suffers from ambiguity.

The FMA has been used in several ontology-based applications (cf. Section 3.4) and, as we shall see in this text, by other biomedical ontologies themselves such as the ontology of ECG we propose in Chapter 5. Being derived from the term-centered UMLS, however, and also due to its large scope, it is in need for several improvements to evolve into a more sound ontological basis.

## 3.2 Biomedical Ontologies' Adherence to Ontology

In line with what has long been claimed in the formal ontology literature (cf. Section 2.2), there is a growing trend in the biomedical ontology literature towards the promotion of principled ontological theories. James Cimino has drawn attention already in (1998) to the desiderata of *non-vagueness* (terms must correspond to at least one meaning), *non-ambiguity* (no more than one meaning) and *non-redundancy* (meanings correspond to no more than one term). However, biomedical terminologies (or even ontologies) are still beset by vagueness, ambiguity and redundancy problems that often lead to inconsistency. More recently, several articles such as (67, 16) have underlined the need for principled biomedical ontologies. In general, it is shared that for biomedical ontologies to be robust enough they ought to hold explicit formal definitions of its relations and universals as much as possible. In order to manifest here that this call for has been widespread, we report three quotations taken from referred articles in what follows.

In an article dealing with logical properties of foundational relations (e.g. *parthood*) (16), Bittner and Donnelly highlight that ambiguity and inconsistency hamper, if not make impossible, the integration of evolving biomedical ontologies to existing terminology systems developed for clinical medicine and bio-medical research. They state that

At least one major obstacle to such integration is that many existing bio-medical terminology systems and ontologies handle foundational relations such as parthood ambiguously and inconsistently. Necessary first steps in overcoming this problem are: (i) to identify the logical properties characterizing specific foundational relations and (ii) to develop a combined

<sup>5</sup>Reasoning over FMA's partonomy is one of the major use cases of the FMA, cf. Section 3.4.

representation of different types of foundational relations in a single deductive system that is expressive enough to make critical distinctions in logical properties explicit.

In (72, p. 107), Schulz et al. in turn draw attention to the need for assigning to each entity a precise formal definition. They sustain that “true definitions” is a requirement rarely met by any existing biomedical ontology. By “true definitions”, they mean (as usual) the description of both the necessary and sufficient conditions for, say, relation instantiation, or class membership. We then refer to a quotation taken from Hahn and Schulz (97).

Basically, all the vocabularies, thesauri, or classifications currently in routine use rest on informal specifications. This means their semantics is rooted in the human (expert-level) understanding of natural language - at least, in the rudimentary form of ‘semantically’ controlled terms and implicit assumptions about the nature of taxonomic, partonomic or otherwise quite unspecific relations between terms (e.g., “related-to” or “associated-with”). Interpreting these relations in the light of a given search or decision support problem or, even more challenging, drawing ad hoc inferences, e.g., following several relations by navigating through a thesaurus often leads to strange or sometimes even bizarre, yet error-prone results. This is usually due to a lack of a rigid, formal semantics underlying these concept systems.

To consolidate the point highlighted here, we conclude with a specially relevant quotation from Ceusters et al. (59).

[...] Mistakes such as these are usually introduced by using relationships that are too generic. Some ontology builders, it is true, adhere to a minimal ontological commitment paradigm, arguing that an ontology should make as few claims as possible about the domain that is being modeled. On our view, however, the job of ontology is not the construction of simplified models; rather, an ontology should correspond to reality itself in a manner that maximizes descriptive adequacy within the constraints of formal rigor and computational usefulness. Terminology authors tend to use relationships that are too generic but giving them (unconsciously) a range of more specific meanings in different types of cases.

Indeed, the ontology background provided in the previous chapter makes the case that formal definitions must be attained whenever is possible. This can be said to be a consensus also in the biomedical ontology community, and we shall draw attention to a somewhat stronger issue. As made by Thomas Bittner in the biomedical ontology community<sup>6</sup>, we echo Guarino (12) once more (cf. Subsection 2.1.3): in order to produce a good ontology one must pursue the *ideal ontology*, i.e., the necessary and sufficient representation of the domain under scrutiny. Moreover, we have seen in Section 2.3 that to do it one needs an ontology formalism that is expressive enough.

Nonetheless, after an overview of some of the most referred biomedical terminologies / ontologies in the literature, we can say that the field of biomedical ontology is *in practice* still far from adhering to its supposed anchoring field, that of Formal Ontology. Besides the quality issues discussed already when introducing the biomedical terminologies / ontologies considered above, one point that turns out a controversy is the way the term ‘reference’ (as in *domain reference ontology*) has been used in the biomedical ontology literature. In an influential article (98), Burgun points to the FMA and ChEBI as two examples of *domain **reference** biomedical ontologies*. In point of fact, Burgun’s definition for *domain reference ontology* - ontologies developed independently of specific objectives (Ibid., p. 307) - is oblivious to a concern with the ontology *accuracy* in fitting into the domain it is supposed to represent. It, rather, only requires such an ontology to be application-independent. Burgun’s definition and her allusion to FMA and ChEBI find support, say, by Bodenreider and Stevens in their briefing on bio-ontologies (13, p. 268). However, those two ontologies can hardly be characterized as reference ones in the spirit conveyed by Guarino (cf. the end of Section 2.3) - which is adopted in the general ontology literature, see e.g. (33). One of the reasons is that they are specified in logical formalisms that fall short in serving as a

<sup>6</sup>Confer slides of Thomas Bittner’s presentation in the Workshop on Ontology and Biomedical Informatics, Rome 2005 (<http://ontology.buffalo.edu/05/wg6/bittner.ppt> - accessed on April 05, 2009).

reference ontology specification framework. As we have seen in Chapter 2, the ontology accuracy rests partially on the expressiveness of the ontology formalism used.

Altogether, it seems like the biomedical ontology community is still beset to move from the legacy of biomedical terminologies to well-founded biomedical ontologies. Nonetheless, we share Bodenreider and Stevens's consideration that "*The successes and, more generally, the developments observed in the field of bio-ontologies over the past 5 years certainly make sense in today's context*" (13, p. 270).

### 3.3 Concept- vs. Realism-orientation

This section is concerned to give a brief account of an ongoing controversy in the biomedical ontology literature about whether *classes* are the extension of cognition-independent types or universals, or of mind-dependent concepts (8, p. 241).

As discussed before, some of the most referred ontology definitions in the Computer Science literature, viz., Gruber's and Guarino's definitions, make use of the notion of "conceptualization". As put by Yu (26, p. 253), Masolo and Guarino et al. refrain from committing to "*a strictly referentialist metaphysics related to the intrinsic nature of the world*". Such a *concept-orientation* has also been stood for in the biomedical ontology literature. For instance, consider Alexa McCray's statement below taken from (68).

It is the thesis of this paper that it is necessarily the case that every conceptualization is biased. This is because representing, or categorizing, the world depends on at least two crucial factors (1) the purpose for which the conceptualization is being created, and (2) the world view of its designer, with the corollary that this depends on the state of general knowledge at the time, as well as on the personal knowledge of the designer.

McCray adheres to the ontology definition of Gruber. Barry Smith, on the other hand, has drawn attention to that the term "concept" is too ambiguous, since it is assigned to at least four different meanings pertaining to *linguistic*, *psychological*, *epistemological* and *ontological* families of view (28, p. 288). In face of this, Smith has proposed a *realism-orientation* (Ibid.) which is defended in depth by Ingvar Johansson in (27). For a brief account, consider the following note assigned by Smith to comment Eugen Wüster's point of view regarding terminologies, which came to be standardized by ISO (75).

Objects may be material (e.g., an engine, a sheet of paper, a diamond), immaterial (e.g., conversion ratio, a project plan) or imagined (e.g., a unicorn) (75). Similarly, Wüster's definition of object would seem to imply that the extension of the concept pneumonia should be allowed to include not only your and my pneumonia but also, for example, cases of unicorn pneumonia or of pneumonia in Russian fiction. With this, I believe, ISO undercuts any view of the relation between concepts and corresponding objects in reality that might be compatible with the needs of empirical science (where it is important to recognize that an imagined mammal is not a special kind of mammal). It thereby also cuts us off from any coherent understanding of that what it is on the side of reality to which the concepts used in biomedicine or other scientific disciplines would correspond.

In our understanding, an important point implicit in the passage above which is made explicit in (27), is not exactly whether or not we suffer from bias, or if should we believe we are able to actually find the truth (independent of human cognition), but rather that *an ontologist must be committed to seek the truth*. To quote Johansson, "*All research needs a regulative idea, something that tells the researchers what to look for. Traditionally, the overarching regulative idea [of both philosophy and science] has been truth*" (Ibid., p. 276). Johansson argues that we should pursue this purpose whether we are supposed to converge in our (accumulative) representations of

the biomedical domain. The object of inquiry along these lines should include empirical evidences, instead of only truth theories or formalisms. The latter still keeps itself in need for metaphysical analysis in order to express our theoretical hypotheses supported by such empirical evidences. However, we would share the regulative idea of seeking the truth as a unity for what we are looking for. For a full discussion, cf. (27), wherein Johansson proposes *Popper's epistemological realism* as a philosophical framework for bioinformaticians, instead of *fictionalism* or the *biasism* suggested by McCray (68).

In the following we carry out an attempt to harmonize Johansson and Smith's realism-orientation to Guarino's cognitivist view. First of all, it is worth to mention that rather than Smith's BFO (52), DOLCE (49) has been applied in domains not strictly concerned to empirical sciences, e.g., services (99), e-Government Ibid. and so on. Such domains are likely even more subjected to human bias. However, the term "conceptualization" in fact seems to let open the following. Since an ontology is *most likely* no more than a human conceptualization - cf. Popper's epistemological realism in (27, p. 278), some researchers tend to take advantage of that to relax the notion of ontology to (just) a shared conceptualization which has not necessarily a commitment to reality. In other words, this trend let ontologies to be seen as consensus theories distant to the truth as a regulative idea. Along these lines, one could (even unconsciously) feel free in developing an ontology meant to represent the truth-for-some-community, instead of the truth in its essence. This seems to be somehow appealing, since it allows to avoid nuance and complication, and mostly, criticism. In case an ontology is just to represent some community's conceptualization, an eventual modeling mistake can perhaps stay hidden more likely. This seems not to be the aspiration of the Formal Ontology in Information Systems' community<sup>7</sup>.

Nevertheless, James Cimino provides a counterargument to Smith (28, p. 289) in defense of the desiderata propounded in (10). He contends that in the biomedical domain itself there is a need for living with both cognition-dependent and -independent entities. Literally from (69), "*I suggest a path that acknowledges the importance of representing reality, as best we can know it, but accepts the need for concepts to help us, among other things, reason under uncertainty. I consider this the **realistic path**. [...] It is my experience that not only can concepts and universals coexist in the same controlled terminology, but that this is a desirable situation.*". One of the examples given by Cimino to enlighten his point is the following.

Consider, for example, "severe acute respiratory syndrome" (SARS). When the condition first arose, we might have chosen to define this term based on a set of actual cases in reality that shared a set of particular attributes (i.e., certain clinical manifestations with particular geographic and chronological characteristics). While such characteristics were certainly true for each individual patient, we must also consider how clinicians dealt with this condition. Did they hold in their minds the unique identifiers of the individual cases or did they use some abstract representation, based on their understanding of the disease at the time? It is certainly the latter, for without the conceptual representation, they would have no way to consider cases that fit some of the pattern of the disease without having all of the characteristics of the initial cases. [...] Smith does not say how we would know to relax those constraints to recognize these cases as being instances of the SARS universal. Humans, however, achieve such reclassification readily, even subconsciously. Instead, clinicians made use of a SARS concept, which included conjectures, such as "probably viral".

Even though without any explicit reference to the discussion above, Schulz et al. (74) state that in what concerns the representation of the biomedical domain there are two quite different objects of analysis, viz., the reality of the patient and all surrounding entities in the health care process on the one hand, and the representation of the health care record, of physicians' knowledge and beliefs on the other hand (100). We then conclude this

<sup>7</sup>See the abstract of the panel "How do we measure progress in formal ontology research?" conducted by Chris Welty in FOIS 2008 (<http://fois08.dfki.de/joomla/index.php/program> - access on April 10, 2009).

section without more considerations, but anticipating that the assumptions behind the development of the ECG Ontology are posed explicitly in Section 5.1.

## 3.4 Applications of Biomedical Ontologies

Despite all the open problems inherent to a young research field, the field of biomedical ontology is bearing witness of relevant results regarding applications. In a relevant article (98), Burgun points out three foremost applications for which a “domain reference ontology” has been used.

1. *Managing heterogeneity of information.* Domain reference ontologies should provide domain knowledge that can be used as a common framework for semantically driven integration of information from different sources that use different terminologies.
2. *Reasoning about complex entities.* Modeling of complex biomedical classes requires knowledge in basic science such as anatomy. For example, the characterization of diseases is based on several relations, including location, which relates disorders to anatomical entities. An ontology of (say) anatomy should provide knowledge in anatomy necessary to perform complex high level reasoning about diseases.
3. *Reasoning about individual data.* Although not designed for specific applications, ontologies must provide generic knowledge for reasoning about individual data in various systems.

In managing heterogeneity of information, for instance, ontologies such as the Foundational Model of Anatomy (FMA) (7) have been used for semantically driven integration. Gennari et al. (101) developed an anatomy platform based on FMA to align Gene Ontology (18) to Mouse Genome Database (102). Their purpose is integrating data sources in genomics to link mouse disease models to human pathological conditions.

In turn, reasoning about complex entities and individual data can be very valuable as well. For example, the Virtual Soldier project<sup>8</sup> comprises the use of both general anatomical knowledge (also based on FMA) and specific computed tomographic images of individual soldiers to aid the rapid diagnosis and treatment of penetrating injuries (103). The patient anatomy is modeled geometrically and each geometric structure is linked to the corresponding anatomic class in the FMA. Reasoning service over FMA’s partonomy are then supposed to predict the damage to organs injured by a projectile, e.g., ischemic regions of the heart when a coronary artery is severed.

Other reasoning-based application is reported by Dameron et al. (104), in the field of Oncology. The application is based on a simplified ontology of lung tumors inspired on the NCI thesaurus (81) combined with reasoning services for automatically grading lung tumors. Grading a tumor consists of matching the description of the location and features of the tumor (and of its possible metastasis) to a grade’s definition. The motivation in this case is that assessing the grade of a tumor can be tough since it requires dealing with different levels of granularity.

The interested reader in a broader list of biomedical ontology applications is referred to (105, p. 88) for more information. In that article, Rubin et al. provide a briefing of biomedical ontologies under a functional view. They intended also to introduce “*those not already highly knowledgeable about biomedical ontology*”, under the feeling that “*it can be helpful to organize ones thinking about them from a functional perspective - how can ontologies be used to enable biomedical research*”.

---

<sup>8</sup>A project from the US Defense Advanced Research Projects Agency (DARPA). Available at <<http://www.virtualsoldier.us/>>.

## 3.5 Conclusions

The key points we have developed throughout this chapter are:

- As it turns out in the literature, the field of biomedical ontology is still beset to move from the legacy of (i) biomedical terminologies (term-centered artifacts) or (ii) biomedical ontologies still too tied to (say) database representation systems, to well-founded biomedical ontologies.
- In virtue of the lack of a correlated work regarding an ontological analysis of the ECG (neither of heart electrophysiology), we have opted to refer to prominent initiatives in the biomedical ontology context in view of setting parameters for a critical analysis of our work.
- In spite of all problems inherent to a young research field, biomedical ontology has seen several applications that (i) exhibit already some of the benefits achieved, and then (ii) justify the efforts which have been made in this field.

As follows we provide an overview of the materials and methods used in the course of our research.

## 4 Materials & Methods

This chapter introduces the materials (Section 4.1) and methods (Section 4.2) used in the research we report in this thesis. We start by introducing the top-level ontology UFO in Subsection 4.1.1, and proceed with a description of the OBO Relation Ontology in Subsection 4.1.2. We then present in Subsection 4.1.3 a brief introduction to the Ontology of Functions. These are the materials used to develop the ontological theory of ECG. Subsequently, we provide a brief description of the combination of semantic web languages OWL DL / SWRL (Subsection 4.1.4), which we have used to specify the ECG ontological theory in a lightweight ontology formalism. Finally, in order to put our methodological choices as clear as possible, we provide a brief description of them.

### 4.1 Materials

#### 4.1.1 Unified Foundational Ontology and OntoUML

The Unified Foundational Ontology (UFO) (53) started as a unification of the GFO (Generalized Formalized Ontology) (50, 51) and the Top-Level ontology of universals underlying OntoClean (45). However, as shown in (53), there are a number of problematic issues related the specific objective of developing general ontological foundations for conceptual modeling which are not covered in a satisfactory manner by existing foundational ontologies such as GFO, DOLCE (48) or OntoClean<sup>1</sup>. For this reason, UFO has been developed into a full-blown reference ontology based on a number of theories from Formal Ontology, Philosophical Logics, Philosophy of Language, Linguistics and Cognitive Psychology. In this way, the development of UFO has been focusing in providing a sound theory for addressing a number of classical conceptual modeling problems.

UFO comprises theories dealing with parts and wholes, types and taxonomic structures, relationships, attributes and attribute value spaces, role playing and qua individuals, among other things (34). Moreover, more recently, this theory has been expanded to deal with dynamic entities such as processes, events and time and with social entities such as action, agent, intentionality, social dependence and delegation, among other things (106). Figure 12 below depicts a small fragment of UFO collecting the ontological categories which will be needed for the ECG domain ontology. In the sequel, these categories are briefly introduced focusing on aspects which are germane to the purposes of this thesis.

The model depicted in Figure 12 shows only types of entities, i.e., from a philosophical standpoint, it is an ontology of universals, not one of particulars. A fundamental distinction in this ontology is the one between *endurants* and *events* (or *perdurants*), which roughly reflects the common-sense distinction between objects (e.g., a car, a person) and processes (e.g., a race, a business transaction). Endurants persist in time maintaining their identity. Events, in contrast, unfold in time with their multiple temporal parts and can be either *atomic* or *complex*,

<sup>1</sup>For consulting the main differences between UFO and OntoClean, the reader can refer to (34, Section 4.5).

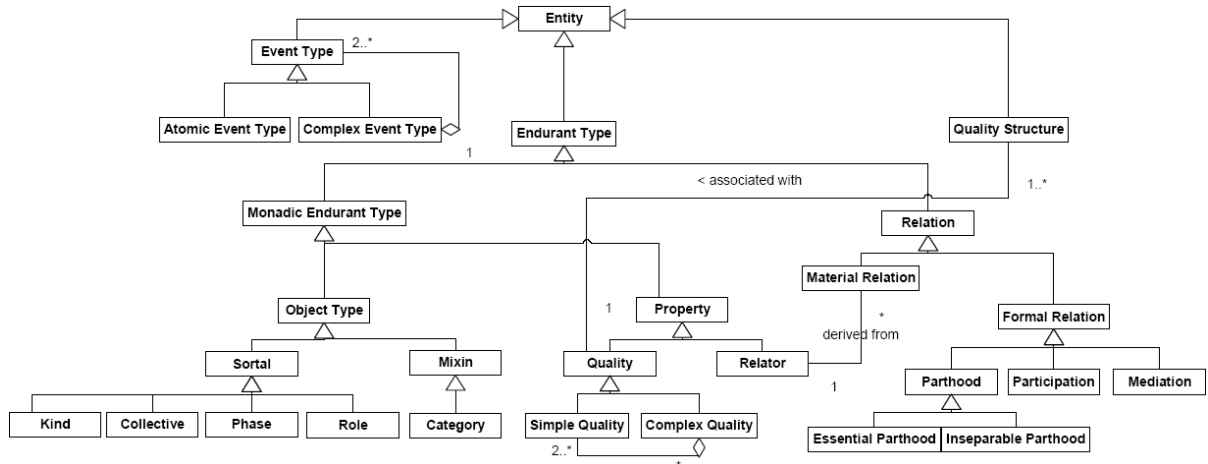


Figure 12: Excerpt of the UFO ontology with elements used in the development of the ECG domain ontology.

the latter being composed of multiple (also possibly complex) events. A formal relation of *participation* is defined between endurants and events. Endurant types can be either types of *monadic entities* or *relations*. Monadic entities, in turn, can be further categorized into *objects* (again, the stereotypical examples in natural language) and *properties*. Property instances are entities which are existentially dependent on other entities, in the way, for example, that the color of an apple depends on the apple in order to exist, just as the symptom of a patient does not exist if not inhering on the latter. A color is an example of property which is projected into some conceptual space, while the symptom exemplifies somewhat not passible to be reduced likewise. Both them are said to be *moments*<sup>2</sup>; this notion comes from the early theory of individual accidents developed by Aristotle in his *Metaphysics* and *Categories*. However, entities like colors are distinguished from those like symptoms under the rubrics of *quality* and *mode*. In conformance with DOLCE (49), UFO distinguishes between the color of a particular apple (its quality) and its value (e.g., a particular shade of red). The latter is named *quale*, and describes the position of an individual quality within a certain *quality dimension*. This notion is an attempt to model the relation between moments and their representation in human cognitive structures carried out by the Swedish philosopher and cognitive scientist Peter Gardenfors which is presented in the theory of *conceptual spaces* (108). For example, while the *quality* height is associated with an one-dimensional structure with a zero point isomorphic to the half-line of nonnegative numbers, color is represented by several dimensions, viz., hue, saturation and brightness.

Property instances can be existentially dependent on single entities (intrinsic moments like qualities and modes) or on multiple entities. These are the relational moments, or *relators*. Examples of the latter include a marriage, a covalent bond, an employment, an enrollment. By the fact that these entities are existentially dependent on multiple entities, they provide the material connection between their depending entities (their bearers). In other words, we can say that they are the foundation for *material relations* such as being married to, being connected to, working at, studying at, etc. Thus, material relations require *relators* in order to be established. *Formal relations*, in contrast, hold directly between individuals. In the ontology of ECG, we consider the formal relations of *parthood*, *participation* and *mediation* (a type of existential dependence). For the relation of parthood we further recognize the case of *essential parthood* - in which a whole cannot exist without that specific part (a case of specific constant existential dependence from the whole to the part); and the case of *inseparable parthood* - in which a part cannot exist without that specific whole (a case of specific constant existential dependence from the part to the whole)

<sup>2</sup>From the German *Momente* in the writings of Husserl. It is also called *trope* in the literature (107).



(109, 110).

While persisting in time, objects can instantiate several object types. Some of these types an object instantiates necessarily (i.e., in every possible situation) and then define (from a metaphysical standpoint) what the object is. Only such substantial sortals can supply a principle of identity for its instances. These are the types named *kind* (for general objects), *collective* (for collections of entities), or *quantities* (for maximally connected entities). There are however types that an object instantiates in some circumstances but not in other circumstances. These are named *phases* and *roles*. While phase is a type instantiated in given time period but not necessarily in all periods and due to the presence of an intrinsic property, a role is a type instantiated in a given context such as the context of a given event participation or a given relation. Examples of the former are the phases **Alive** and **Deceased** of the kind **Person** motivated by the presence or not of an intrinsic property of being alive. An example of the latter is the role **Patient** again instantiated by instances of the kind **Person** and motivated by the presence or not of relational property of being treated in a given medical unit. Finally, a *category* is a type that classifies entities that belong to different kinds but that share a common essential property (i.e., a property that they must not lack). As such, a category can not have direct instances, which are rather instantiated directly by kinds or collectives. For example, the category **Rational entity** subsumes the disjoint kinds **Person** and **Artificial agent**.

All this have been introduced here with the solely purpose of providing an overview. We provide, notwithstanding, a more detailed explanation of those meta-categories by demand as they are used in our ECG ontological theory. For a complete definition and in-depth discussion about these categories, please refer to (34). In (34), the ontological distinctions just described (among many others) are formally characterized using a system of quantified intensional logic with sortal-restricted quantification. The features of this logic and its corresponding instantiation for the ECG domain ontology cannot be elaborated on here. Nonetheless, in the sequel we briefly illustrate some modal constraints which are implied by the use of some of the elements defined in this foundational ontology. For instance, consider the characterization of the essential parthood (EP) relation. This is expressed in formula schema P1 below such that: a universal  $A$  has an essential parthood relation with a universal  $B$  iff for every instance  $x$  of  $A$ , there exists a specific instance  $y$  of  $B$  such that in every situation (possible world) that  $x$  exists,  $y$  also exists and is a *part of*  $x$  in that situation.

$$(P1) \quad \mathbf{EP}(A, B) =_{def} \Box (\forall x \mathbf{A}(x) \rightarrow \exists y \mathbf{B}(y) \wedge \Box (\varepsilon(x) \rightarrow \mathbf{part\_of}(y, x)^3))$$

One real world case in which the essential parthood can be illustrated is the relation between a *car* and its *chassis*. In every possible world  $w$  in which a car exists, his chassis must also exist and it must be part of the car, and this particular part (chassis) cannot be changed, i.e., cannot be replaced by another part of the same type. Inseparable parthood works likewise, but with predicate  $A$  denoting the part and  $B$  the whole. An example can be given by the parthood relationship between an alive (human) brain and an alive (human) body. The former cannot materially exist if it is not part of the latter for its whole lifecycle, if we assume not brain transplant to be possible (109).

Another example of the UFO axiomatization can be given by considering the rigidity property inherent to some of the object types, namely the *kinds*, *collectives* and *quantities*. Rigidity is formally described by  $P2$  to capture that: a universal  $A$  is rigid iff for every instance  $x$  of  $A$ ,  $x$  is necessarily (in the modal sense) an instance of  $A$ . In other words, if  $x$  instantiates  $A$  in a given world  $w$ , then  $x$  must instantiate  $A$  in every possible world  $w'$ .

<sup>3</sup>The predicate labeled with the Greek symbol  $\varepsilon$  denotes material existence. It is adopted in this axiomatization what is termed mereological continuism which states that parthood relations can only hold between existing entities. This renders the inclusion of the predicate  $\varepsilon(y)$  in the consequent of formula (P1) superfluous.

$$(P2) \quad \Box ( \forall x A(x) \rightarrow \Box A(x) )$$

Rigidity is one of the meta-properties firstly used in Computer Science by Guarino and Welty (46) to characterize the elements of an ontology. For clarity, let us also consider the anti-rigidity property, which is used to define object types such as roles and phases. Anti-rigidity is formally characterized by *P3*, in saying that: a universal *A* is anti-rigid iff, for every instance *x* of *A*, *x* is possibly (in the modal sense) not an instance of *A*. In other words, if *x* instantiates *A* in a given world *w*, then there is a possible world *w'* in which *x* does not instantiate *A*.

$$(P3) \quad \Box ( \forall x A(x) \rightarrow \Diamond \neg A(x) )$$

The universal **Person** is an example of a rigid object type, whereas the universals **Student** and **Living person** are examples of anti-rigid object types (role and phase, respectively). In the following chapter, the usefulness of such ontological distinctions is evidenced in the representation of the ECG domain. However, it is still worthwhile to introduce before how those UFO ontological categories have been used by Guizzardi et al. to develop an ontologically well-founded UML profile named **OntoUML** (111).

The modeling primitives of that profile directly stand for the ontological distinctions postulated by UFO. Moreover, the axiomatization prescribed by the foundational ontology and which governs the admissible ways in which ontological categories can be combined is incorporated in the language meta-model as formal constraints. As a result, the language “deems” (by means of its grammar) as grammatically incorrect all specifications produced which are non-conformant with the axiomatization prescribed by UFO. Rigidity characterizes in OntoUML, for instance, the stereotype *kind*, since the *kind* universal is one the UFO categories which bears the meta-property of rigidity and then stands for the axiomatization described by formula P1.

In other words, the use of such a language offers not only the feature of making explicit the ontological categories comprising a domain representation, but it also restricts the construction of models to only those which are ontologically consistent. Consequently, ontological engineering can benefit from that in terms of providing the modeler with a support for constructing ontologically well-founded conceptual models. Benevides and Guizzardi have developed a tool whose functionality is to parse conceptual models specified in OntoUML and let the user know about the ontological flaws (112). The benefits aforementioned are what we have tried to expose with the discussion laid in Section 2.3 and the example given in Fig. 5.

In Chapter 5, we use OntoUML in addition to FOL to convey our ECG ontological theory. As an example of OntoUML primitives, we can cite the essential and inseparable parthood distinctions formally described above. They are represented in OntoUML by means of the *essential=true* and *inseparable=true* tagged values added to the usual lines denoting part-whole relationships in UML. The UFO meta-categories introduced above like *kind*, *role*, *mode* and so forth are likewise represented by stereotypes as *kind*, *role* and *mode* that distinguish the usual and overloaded UML class. In this way, the UFO ontological meta-categories which are *instantiated by* domain universals are made explicit.

At this point it may be worthwhile to draw attention to reason why the relationship between UFO meta-categories and domain entities are that of instantiation, and not of subsumption. This is often confused such that the relationship just mentioned is assigned to be subsumption. As elaborated by Guarino and Welty in (44, p. 64), however, “*subsumption is not meta*”. The somehow appealing assertion that a class “Rigid class” subsumes all classes that are rigid such as **Human** is an example of subsumption misuse. The reason is that

“...a quick look at identity criteria reveals that this relationship cannot be. Instances of “rigid class” are classes, which can be identified in various ways (intensionally, in terms of the properties that define the class, or extensionally, in terms of their members). In any case, these identity criteria cannot be applied to the instances of Human, so being rigid is a meta-property of the class Human and *subsumption is not meta*”.

Accordingly, the class *Person* is not a *subtype of kind*, but rather an *instance of* such. An additional illustration for that can be given by considering some instance of *Person*, say, me. If *Person* were a subtype of *kind* (or more generally, of a class *Rigid class*), *Student* would have to be likewise since it is a subtype of *Person*, cf. subsumption’s (or *is-a*) definition in Table 1. However, *Student* is anti-rigid<sup>4</sup>, since I am an instance of *Student* at the time of writing, but will be no longer after I have finished the Master’s in Informatics at UFES, cf. P3 above. We then reach a contradiction.

### 4.1.2 OBO Relation Ontology

The OBO Relation Ontology (RO) (67) has been developed to enhance the treatment of relations in biomedical ontologies. It provides a methodology for providing consistent and unambiguous formal definitions of the relational expressions used in biomedical ontologies. The RO has been developed through a collaboration between formal ontologists and biologists in the OBO, FMA and GALEN research groups and also incorporates suggestions from a number of other authors and curators of biomedical ontologies. Due to its generality in the biomedical domain, the RO is expected to promote interoperability of biomedical ontologies and support new types of automated reasoning over biomedical entities (67).

The RO in fact standardizes basic relations that cross-cut the biomedical domain. Most of them deal with spatial and temporal aspects of biomedical entities. They therefore can be used in the definition of domain-specific relations (e.g., the conduction of the cardiac electrical impulse). A fundamental distinction in many foundational ontologies (e.g., DOLCE, BFO, GFO, UFO) and, in particular, in the OBO Relation Ontology (RO), is the distinction between (in a loose sense) objects and processes. In BFO (52), for example, these notions are named continuants and occurrents, respectively. To put it literally from (67),

Continuants are those entities which endure, or continue to exist, through time while undergoing different sorts of changes, including changes of place. Processes are entities that unfold themselves in successive temporal phases.

Generally speaking, the notion of continuant can be said similar to what is called *endurant* in UFO, while process can be seen similarly as a *perdurant*. Table 1 presents the RO relations which we employ here. A full discussion on these relations can be found in (67). Initially, we keep here the semi-formal syntax employed in that article to further move on to their corresponding First-Order Logic (FOL) counterparts. The following variables and ranges are used in the sequel.

$C, C_1, \dots$  to range over continuant classes

$P, P_1, \dots$  to range over process classes

$c, c_1, \dots$  to range over continuant instances

$p, p_1, \dots$  to range over process instances

$r, r_1, \dots$  to range over three-dimensional spatial regions

<sup>4</sup>According to a widespread view in the Conceptual Modeling literature (113).

$t, t_1, \dots$  to range over instants of time

Table 1: The relations of the OBO Relation Ontology that we make use in this thesis.

<i>Relation</i>	<i>Definition</i>
$c$ <b>instance_of</b> $C$ <b>at</b> $t$	a primitive relation between a continuant instance and a class which it instantiates at a specific time
$p$ <b>instance_of</b> $P$	a primitive relation between a process instance and a class which it instantiates holding independently of time
$c$ <b>part_of</b> $c_1$ <b>at</b> $t$	a primitive relation between two continuant instances and a time at which the one is part of the other
$c$ <b>located_in</b> $r$ <b>at</b> $t$	a primitive relation between a continuant instance, a spatial region which it occupies at a specific time
$r_1$ <b>part_of</b> $r_2$	a primitive relation of parthood, holding independently of time (i.e., holding constantly) between spatial regions (one a sub-region of the other)
$r$ <b>adjacent_to</b> $r_1$	a primitive relation of proximity between two disjoint spatial regions
$t_1$ <b>earlier</b> $t_2$	a primitive relation between two times
$p$ <b>has_participant</b> $c$ <b>at</b> $t$	a primitive relation between a process and a continuant at a specific time
$p$ <b>has_agent</b> $c$ <b>at</b> $t$	a primitive relation between a process, a continuant at a specific time $t$ at which the continuant is causally active in the process
$C$ <b>is_a</b> $C_1$	for all $c, t$ , if $c$ <b>instance_of</b> $C$ <b>at</b> $t$ then $c$ <b>instance_of</b> $C_1$ <b>at</b> $t$
$c$ <b>exists_at</b> $t$	for some $p$ , $p$ <b>has_participant</b> $c$ <b>at</b> $t$
$p$ <b>occurring_at</b> $t$	for some $c$ , $p$ <b>has_participant</b> $c$ <b>at</b> $t$
$t$ <b>first_instant</b> $p$	$p$ <b>occurring_at</b> $t$ and for all $t_i$ , if $t_i$ <b>earlier</b> $t$ then not $p$ <b>occurring_at</b> $t$
$t$ <b>last_instant</b> $p$	$p$ <b>occurring_at</b> $t$ and for all $t_i$ , if $t$ <b>earlier</b> $t_i$ then not $p$ <b>occurring_at</b> $t$
$c$ <b>located_in</b> $c_1$ <b>at</b> $t$	for some $r, r_1$ ( $c$ <b>located_in</b> $r$ <b>at</b> $t$ and $c_1$ <b>located_in</b> $r_1$ <b>at</b> $t$ and $r$ <b>part_of</b> $r_1$ )

### 4.1.3 Ontology of Functions

The Ontology of Functions (OF) (4, 114) has been developed under the umbrella of the General Formal Ontology (GFO) research program (50). It provides a top-level ontological framework for representing functional knowledge in the biological domain. This framework is used to define and specify functions, and relate them to each other as well as to other entities in Biology. It has been intended to increase the accuracy and expressiveness of biomedical ontologies by providing a means to capture existing functional knowledge in a more formal manner.

In the application of OF in a natural sciences domain such as the heart electrophysiology, a function is considered a teleological entity. That is, somewhat exhibiting or relating to design (or purpose) in nature. The OF addresses three major issues concerning functions (4):

- *function structure*: how to represent and determine functions independently of their realizations;
- *realization*: the conditions under which a given entity realizes a function;

- *has-function relation*: the determination of the notion of an entity having a function.

The basic structure of a function, as introduced in (Ibid.), is a set of labels, a functional item, a set of requirements to be fulfilled in case the function is realizable, and a goal to be satisfied in case the function is in fact realized. A function is connected to a continuant which has the function, and can realize it by playing a specific role (the functional item). This role is exercised by what is named in Philosophy a *qua individual* (115). For instance, if John marries Mary, a number of rights and duties (legally speaking) are to henceforth be satisfied by John-qua-husband-of-Mary. The notion of *qua individual* finds place in the OF according to the theory of roles proposed by Frank Loebe in Ibid.. Otherwise, we take the notion of role into account in this thesis in the terms outlined by Guizzardi in (34, Chapter 7), (113). The basic difference between them is that, in Loebe's assignment, the individual which plays a role, which is subject of action, is a *qua individual* that is existentially dependent on some real individual, but a different one. For example, John-qua-husband-of-Mary (which plays the role of Mary's husband) is not, in this sense, the same individual as John. In Guizzardi's point of view, otherwise, John-as-husband-of-Mary (which plays the role of Mary's husband) and John are the same individual. Those two accounts of roles / *qua individuals*, however, can be theoretically harmonized - cf. Ibid., Chapter 7 - such that the differences does not have any practical implication for the purpose of the theory developed in this thesis.

Furthermore, a function is realized by means of a process (4). This process provides a transition from the state of the world (SOW) in which the requirements of the function are fulfilled, to the SOW in which the goal of the function is satisfied. This process is called the realization of the function. A realization can be considered *actual* or *dispositional*. That is, a process can have the disposition of being the realization of the function, even if this disposition is never actualized, e.g., in case of some malfunctioning. While the *dispositional realization* is somewhat that exists dispositionally in a thing, a certain power or potential which is the product of evolution or design, the *actual realization* is something that takes place episodically, which is the product of intentionality or local causal influence. We shall adopt in this thesis a convention that "realization" and "dispositional realization" coincides. A "realization" however, can be further said "actual" in case there is evidence that the function has in fact been *actually realized* by the referred process.

Figure 13 illustrates two examples of biological functions represented by using the OF framework, viz., to transport sugar and to accumulate oxygen. While the former is realized by means of the process of carbohydrate transport, the latter is realized by means of the process oxygen accumulation. Their applicability is enlightened in (4) by taking into consideration the Gene Ontology and Chemical Entities of Biological Interest (ChEBI).

Furthermore, some characteristic relations are held between functions (4).

- *Supports*: one function supports the other if its goal fulfills partially the second function's requirements (the goal of the first function is a proper part of the requirements of the second function).
- *Enables*: one function enables the other if its goal fulfills all of the second function's requirements (the requirements of the second function are a part of the goal of the first function).
- *Prevents*: one function prevents the other if its goal excludes the requirements of the second.

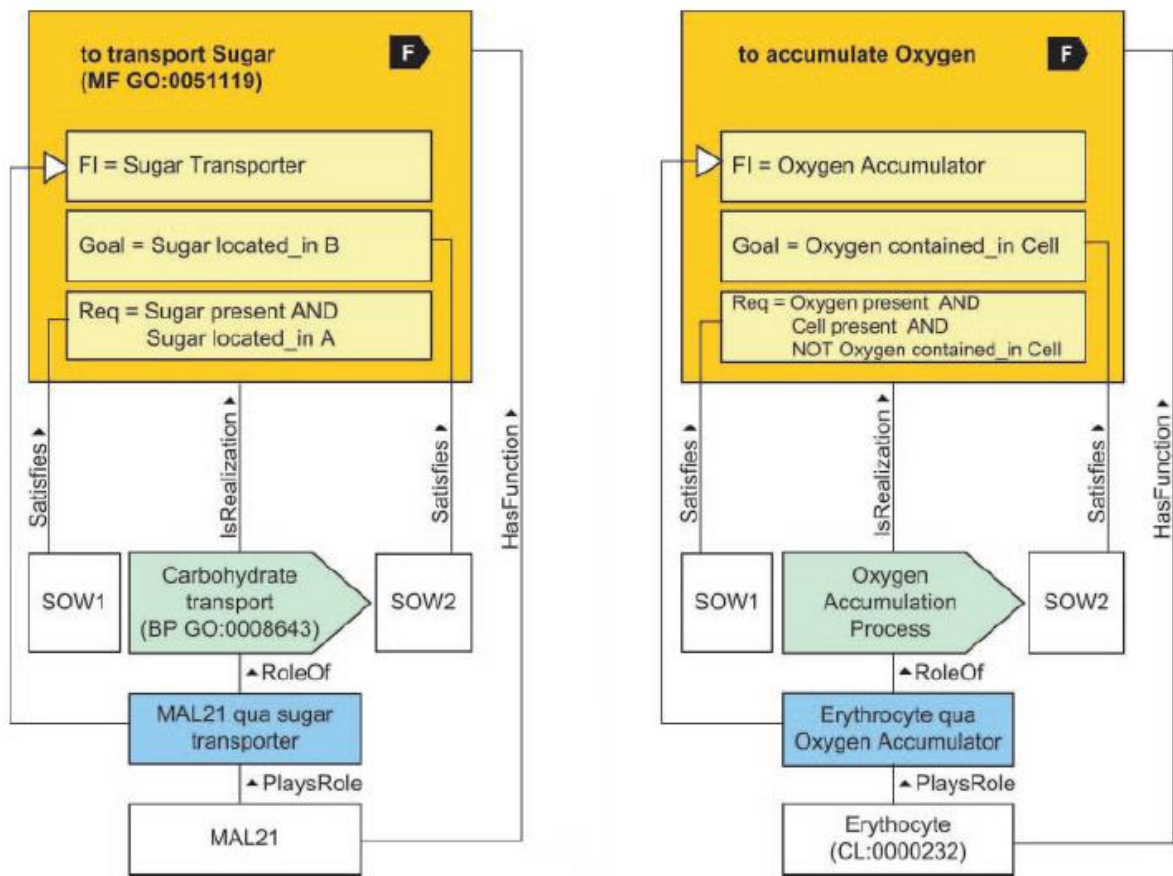


Figure 13: Two exemplary models employing OF (source: (4)).

#### 4.1.4 OWL DL / SWRL

The Web Ontology Language - OWL (116, 57) is a formalism designed to meet requirements for knowledge and data representation on the web. As the OWL language has been elaborated, the Semantic Web Rule Language - SWRL (117, 118) was created to be a rule language built on top of OWL. It was intended to extend OWL for allowing the inclusion of horn-like rules in web ontologies. Both OWL and SWRL are W3C recommendations that constitute noteworthy technologies in the semantic web effort.

For meeting conflicting requirements of the semantic web initiative, OWL was divided into three families. OWL DL is an OWL family based on the Description Logic SHOIN(D), which is strictly designed to guarantee decidability (57); as combined with SWRL, however, in general the automated reasoning procedures turn to be undecidable. In face of this, recent research efforts have provided strategies to overcome this issue. A notion that has been ever more adopted by off-the-shelf semantic web reasoners is that of DL-safety for SWRL rules. It imposes the restriction that the (universally quantified) variables occurring in the body of a SWRL rule must be bound to range over known individuals in the knowledge base (119). This restriction makes OWL DL / SWRL-based reasoning to be decidable, and has shown effective practical results.

In recent years we have seen the emergence of OWL as a *de-facto* standard for knowledge and data representation on the web. Although the combination of OWL DL and SWRL has serious limitations in terms of expressiveness, it bears features for advancing applications on the web. First of all, being a W3C recommendation is a step forward in fostering interoperability between web applications. Second, this language allows one to keep

the domain representation and its instantiation (both the T-Box and the A-Box, in the Knowledge Representation jargon) aside from application code. The same lightweight ontology could then be reused by different applications. Indeed, reuse is also favored in OWL DL / SWRL with respect to the evolution of a given ontology. The reason is that the epistemic treatment of queries (asking about what is known) in OWL is designed to mirror the open-world assumption (120). That is, what is not explicitly stated in the knowledge base is not considered false, leaving then the unknown open to further extensions. Finally, OWL DL / SWRL is designed to afford automated reasoning for deriving new information from existing information in the knowledge base. Speaking of the web of today, this feature is not possible by using, say, XML, which is currently the most likely used language to data representation on the web. Such a reasoning feature is also useful for checking logical consistency and reclassifying classes and individuals at run-time.

The OWL concrete syntax is XML-based, and as such it is a bit verbose. However, as part of the Semantic Web effort, editing tools for building OWL ontologies have been developed. The user who is to build an OWL ontology then has not to deal directly with the OWL concrete syntax, but rather with a much more readable abstract syntax. Figure 14 depicts a screenshot of an OWL ontology's edition in Protégé 3.1 (121), a referred edition tool for building OWL ontologies. Figure 15 in turn shows a screenshot of the SWRL tab, an environment to edit SWRL rules in Protégé. For an in-depth introduction to OWL, the reader may refer to (116) or (122).

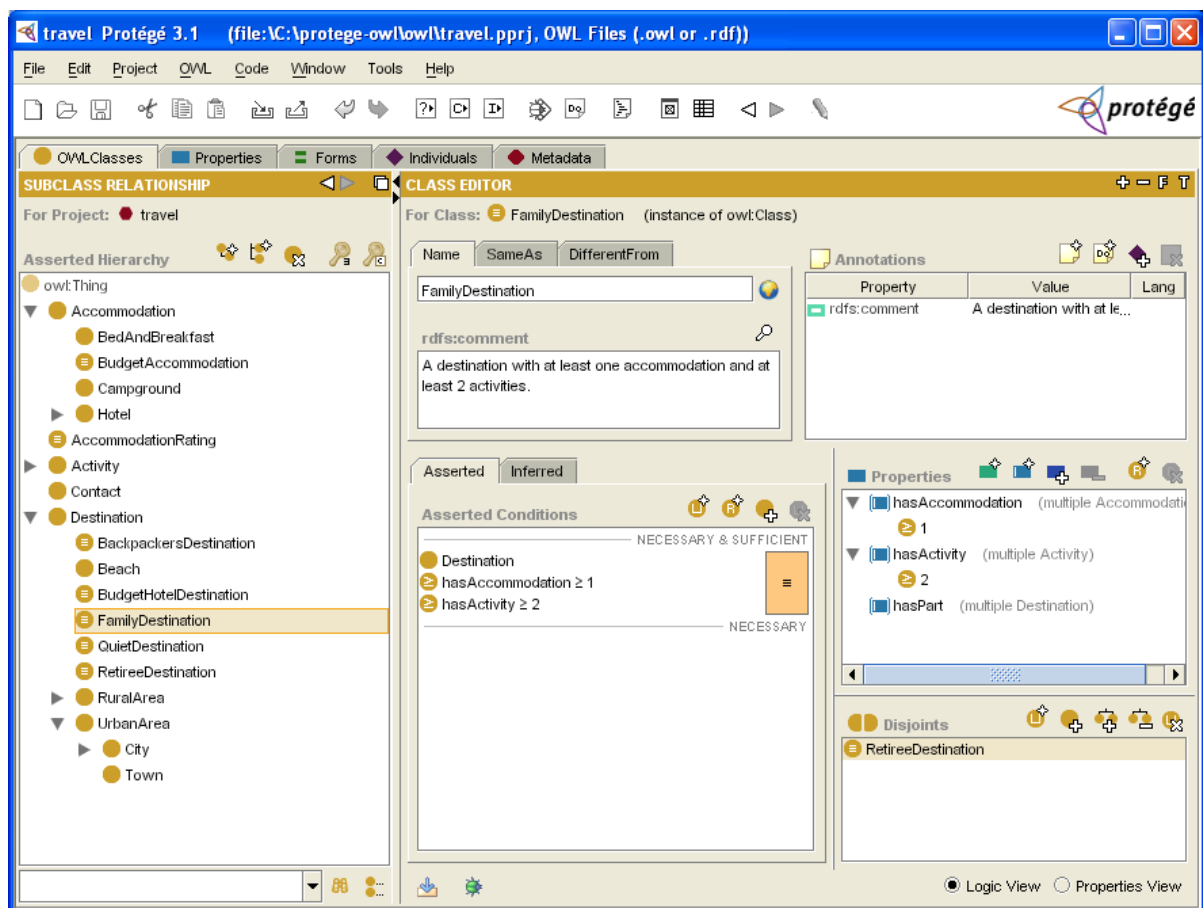


Figure 14: The OWL Classes view can be used to edit hierarchies of classes. Details of the selected class are shown in the right part of the screen. The upper part of this area allows users to add comments, labels and other annotations. The lower part displays logical characteristics of the selected class (source: (121)).

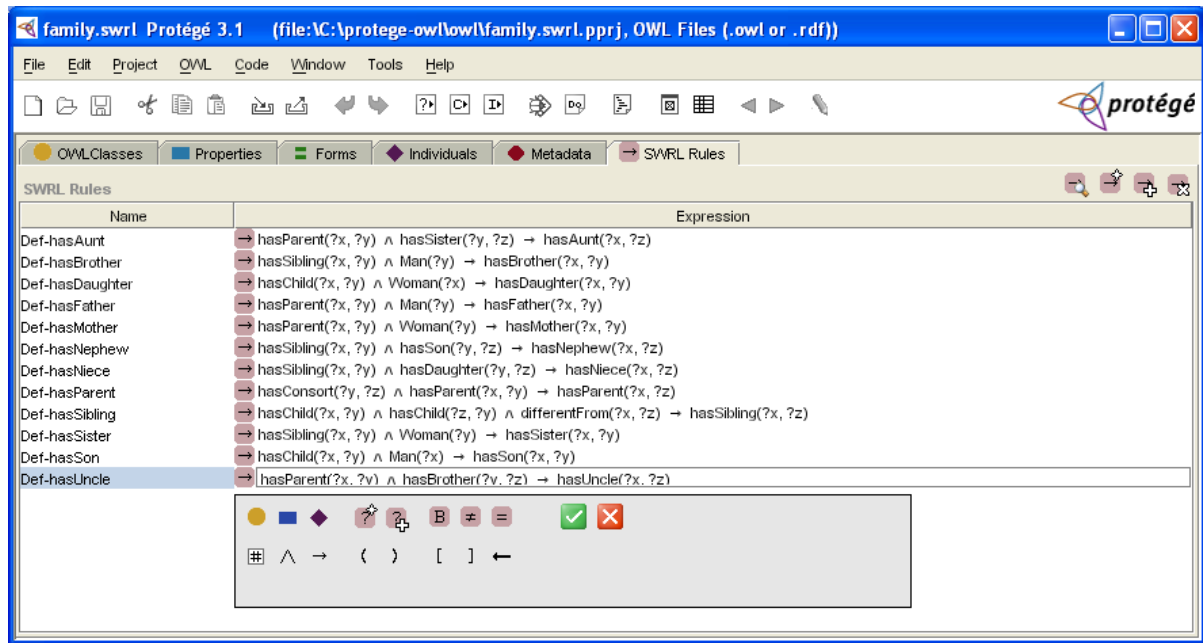


Figure 15: Protégé also supports editing rule bases in the Semantic Web Rule Language (SWRL). Rules can be edited with a convenient expression editor. (source: (121)).

## 4.2 Methods

- Ontological Foundations:** we draw attention to the fact that building a (biomedical) domain ontology on the basis of some ontological foundation is beneficial, if not necessary. In this respect, a top-level ontological framework (cf. Subsection 2.2.2) not only provides us with a support in making ontological decisions (cf. Subsection 2.2.1), but also allows us making these decisions as transparent as possible in the resulting domain ontology. Our ECG ontological study is grounded in the top-level framework of UFO (cf. Subsection 4.1.1).
- Ontology Engineering:** we employ an ontology engineering approach built upon the assumption that fostering a domain ontology (in the AI context) calls for two different ontology artifacts (61), viz., one ontologically well-founded theory of the subject domain meant to be strongly axiomatized for constraining as much as possible the theory's intended meaning, also called *reference ontology*; and another meant to be a computable artifact for automated reasoning and information retrieval, known as *lightweight ontology*. In addition, as traditional in Ontology Engineering (cf. Section 2.4), we specify a set of competence questions for delimiting the scope and purpose of the domain we have at hand. This methodological technique is also beneficial in the end of the development cycle as a means for evaluating the resulting artifact.
- Ontology Formalisms:** according to the engineering approach we have adopted, two ontology formalisms are used to produce the reference and lightweight versions of the domain ontology to be produced. As said before (cf. Section 2.3), Bittner and Donnelly put forward an analogous line of argument and propose the use of First-order Logic (FOL) as a formalism for the former while some sort of Description Logic (DL) could be used for the latter (16). We follow the same choice of ontology formalisms here, particularly by using for the latter OWL DL (mirrored to a DL) with the SWRL extension for rules (118). However, the ECG ontological theory (or reference ontology) is composed by models specified in OntoUML as well(cf. the end of Section 4.1.1). This language is an ontologically well-founded UML profile that brings in the



feature of making the UFO ontological categories instantiated by the ECG domain elements explicit in the ontology.

- **Ensuring effective automated reasoning:** one of the main practical objectives of the research reported here is to use the results of our ECG ontological study for supporting automated reasoning over universals and particulars of ECG and heart electrophysiology. We have then been (practically) pursuing the sweet spot in expressing as much as possible of the ECG ontological theory we develop here in a combination of OWL DL and its SWRL extension, while keeping computational decidability and tractability. Since higher-order logics jeopardize the goal of practical automated reasoning, UFO categories are expressed in the resulting ECG ontology implementation merely as OWL annotations. In spite of this, we contend that the principled structure of the ontology (e.g. the ontological soundness of subsumption and parthood taxonomies) is still preserved in the implementation.
- **Ontology Integration:** we seek ontology integration, specially towards the OBO foundry and the semantic web effort. The latter has influenced us to select the OWL DL / SWRL combination as an ontology codification language. Regarding the former, firstly, our ECG ontological theory has been inspired on the FMA (cf. Subsection 3.1.5) for covering anatomical entities which are relevant for an ECG theory. Besides that, we apply the OF (cf. Subsection 4.1.3) as a framework to model heart electrophysiological functions. Basically, we aim at providing a clear structure of heart electrophysiological functions (what they are), and how and by whom they can be realized. We intend, by these means, to be able to reconstruct those physiological entities from a particular ECG. Secondly, we borrow relations from the OBO Relation Ontology (cf. Subsection 4.1.2) which are especially valuable for defining spatial relations over time. We use them in combination with domain-specific relations coined here and complementary relations formally described in UFO.

At this point, once the basis of our work have been established, we can proceed to present the results achieved in the chapters that follow.

## 5 The ECG Ontological Theory

In this chapter we put the ECG domain under ontological analysis. The analysis is grounded on the Unified Foundational Ontology (UFO), cf. Section 4.1.1. Our very goal here is to inquiry *what the ECG is in essence*, on both sides of the patient and of the physician. With that purpose, the phenomena underlying this cardiological exam are also in need of consideration. We, thus, deal with the domains of human heart electrophysiology and anatomy as well.

It is worth to recall that the ECG theory developed here makes use of a number of existing foundational theories, namely: (i) the OBO Relation Ontology (Section 4.1.2), for defining our domain-specific relations involving space and time; (ii) the Foundational Model of Anatomy, for handling the necessary human anatomy for the ECG; (iii) the Ontology of Functions, for tackling heart electrophysiological functions. Besides, in order to convey the ECG ontological theory, we adopt an ontology formalism that is twofold: (i) OntoUML, an ontologically well-founded UML profile, and (ii) First-Order Logic (FOL) - with identity - formulae.

This chapter is structured as follows. First of all, in Section 5.1 we discuss some conventions and epistemological assumptions underlying the ECG theory we propose in this thesis. We then start to develop the ECG theory in Section 5.2 by taking into consideration the anatomical elements of the human body which are involved in the acquisition of the ECG. Subsequently, in Section 5.3 we turn our study to the human heart electrophysiology that is mapped in the ECG. We then finally reach the ECG itself in Section 5.4, by considering it on both sides of the patient (who is the subject of the exam) and of the physician (who analyzes it). As follows a basic ECG interpretation is introduced (Section 5.5). In Section 5.6, the connection between the ECG and its underlying phenomena are then put under analysis. The ECG Ontology is finally outlined in Section 5.7, which addresses a full documentation of the ECG theory developed in the course of the previous sections. We then conclude the chapter by summarizing the key points developed throughout it.

### 5.1 Preliminaries: Conventions & Epistemological Assumptions

With transparency in mind, it may be worth to list the following points regarding our epistemological assumptions and conventions:

- The ECG theory proposed here accounts for a descriptive commonsensical view of reality, focused on structural (as opposed to dynamic) aspects of the ECG.
- The basic material sources regarding the ECG, the heart electrophysiology and anatomy taken as references here constitute medical textbooks written by authoritative physicians in the field. We assume that the knowledge it contains has been collected by following the scientific method.

- In line with the philosophical considerations made by Ingsvar Johansson in (27), we have carried out the ECG domain representation by looking *through* terms to reach their referents in reality. The latter are our units of representation here. To echo Johansson's optical metaphor distinguishing looking *at* and looking *through*<sup>1</sup>, we aim to use terms *just as an astronomer uses a telescope to see the planets and stars*.
- Although we seek an accurate representation of what the ECG is<sup>2</sup>, we are aware that this will hardly be the case, cf. also Section 3.3. Popper's epistemological realism proposes "*Seek truth but expect to find truthlikeness*". In this sense, we are committed to propose here an evolving ECG theory.
- We assume an Aristotelian view of universals, viz., that they exist in their instances (exist *in re*). For this reason every definition carried out here is firstly made in the instance-level, and posteriorly extended into the universal-level (or class-level).
- From then on in this text the term 'universal' is considered synonymous with the terms 'class' and 'type'<sup>3</sup>. The term 'particular' in turn is considered here as synonymous with 'instance' and 'individual'. Like told by Schulz et al. in (70), this bridges philosophical and computer science jargons, without a concern with the subtleties involving their exact meanings. The distinction between universals and particulars is made explicit by strict conventions: domain universals are introduced in **sans serif** typeface with their names using Upper Case initials, while names of particulars are written in lower case letters.
- Time arguments are used in logical sentences only if necessary. That is, if omitted, then the assertion at hand is time invariant. For example, *Heart*( $c_1$ ) or *CardiacElectricalImpulse*( $c_2$ ) mean that individuals  $c_1$  and  $c_2$  are *instances of Heart* and *Cardiac electrical impulse* in their whole existence - see def. below, while *SANodeMyocytesPolarized*( $c, t_1$ ) means that individual  $c$  is *instance of SA node myocytes polarized* at time instant  $t_1$  - see Table 1 in Section 4.1.2.

The following definition holds for capturing assertions that a certain individual is *instance of* a given continuant for its whole lifetime.

$$\mathbf{instance\_of}(c, C) =_{def} \forall t ( \mathbf{exists}(c, t) \rightarrow \mathbf{instance\_of}(c, C, t) )$$

Notice in the formula above that we did not include in its antecedent the condition that  $\exists t \mathbf{exists}(c, t)$  for avoiding its trivialization - as  $c$  could never actually exist. This is because the conventions we set as follows cover this, but also other issues. These conventions apply for all over this chapter.

First of all, we submit that all the individuals of our interest exist eventually but not necessarily. That is, in this theory we are not interested in individuals that exist necessarily nor in those that necessarily do not exist. So,

$$\forall c \exists t_1, t_2 \mathbf{exists}(c, t_1) \wedge \neg \mathbf{exists}(c, t_2) \wedge (t_1 \neq t_2)$$

Besides, all the universals of our interest are *informative*, i.e., they must be eventually instantiated. This is warranted by the following schema. Let  $C$  be any universal, then  $\exists c, t \mathbf{instance\_of}(c, C, t)$ .

<sup>1</sup>Similar to the traditional distinction between the *use* and *mention* of linguistic entities. For an in-depth reading, please refer to (27).

<sup>2</sup>In Subsection 2.1.3 the notion of accurate representation of a given domain is enlightened. Further on in this chapter, specifically in Subsection 5.7.1, we define our subject domain in an objective way by means of competence questions described in FOL.

<sup>3</sup>We refrain from using the term 'concept' in virtue of its multiple and partially contradictory meanings. We rule out in this thesis the 'class' reading related to a *extensionalist* spirit, i.e., a reduction of intension to set-theoretic and therefore extensional entities (54, p. 35).

Finally, we assume by *actualism*, that  $\text{instance\_of}(c, C, t) \rightarrow \text{exists}(c, t)$ . And by *mereological continuism*, that  $\forall c_1, c_2, t (\text{part\_of}(c_1, c_2, t) \rightarrow \text{exists}(c_1, t) \wedge \text{exists}(c_2, t))$ .

## 5.2 Anatomy for the ECG

In this section we provide an ontological account of the human body anatomical continuants directly involved in the ECG. We take the FMA (94, 7) as a reference, and then consider continuant universals either by using the same terms employed by the FMA or their synonyms. Nonetheless, we have not followed strictly the FMA modeling choices, since they are not fully supported by ontological foundations, cf. Section 3.1.5. In addition, we have also taken Weinhaus and Roberts' chapter on the heart anatomy (123) as a medical textbook reference. Besides, we can say we have targeted a suitable balance in, on the one hand, not to include here universals which are not relevant in the ECG domain, and on the other hand, to be inclusive enough to succeed in clarifying the differentiae used for defining the genus / species hierarchy.

In consonance with the FMA, we consider **Anatomical Entity** to be the most general universal, i.e., our supreme anatomical genus. We consider it to be specialized into the disjoint categories *material* and *immaterial anatomical entities* (see Figure 16). The formal meaning of the *is-a* relationship is given in Table 1.

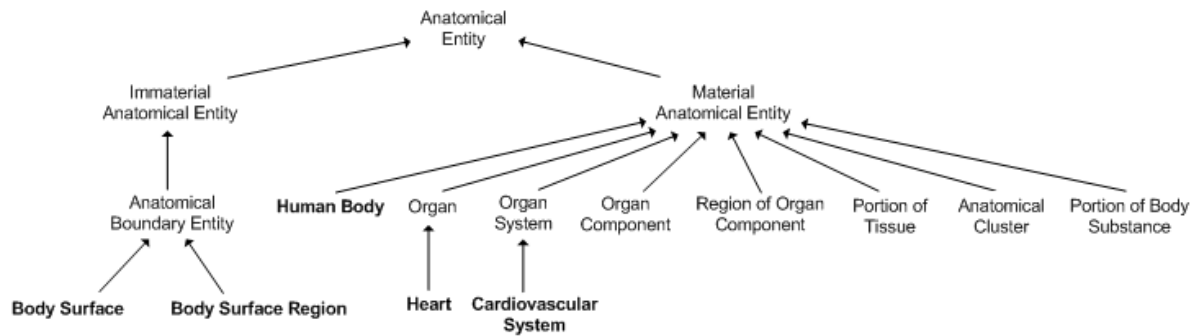


Figure 16: Anatomical entity and its partition into material and immaterial entities, inspired in the FMA. The arrows represent *is-a* relationships from the subtype towards the supertype universal. The universals in boldface are *kinds*; all the others are *categories*. The five categories on the right-hand (viz., Organ component, Region of organ component, Portion of tissue, Anatomical cluster and Portion of body substance) are to be further elaborated.

First, let us focus on the branch of immaterial anatomical entities. In the scope of anatomy that crosscuts the ECG domain, a germane notion is that of **Anatomical boundary entity**. This is because, to anticipate Section 5.4, in order to acquire the ECG, at least two electrodes have to be placed on the **Body surface**. Actually, each electrode is to be placed on a specific **Body surface region** (e.g., the surface of the right wrist). Both of the two universals just mentioned are subsumed by the *anatomical boundary entity* universal. The **Body surface** is something Smith terms a *bona fide boundary*, as opposed to the so-called *fiat boundary* (124). Basically, a fiat<sup>4</sup> boundary is a boundary that exists only in virtue of the different sorts of demarcations effected cognitively by human beings. Genuine, or bona fide boundaries, on the other hand, are all other boundaries, i.e., those that are independent of human fiat (Ibid., p. 5). Accordingly, a fiat object is defined by means of the human process of tracing an arbitrary demarcation to shape it. Broadly, it is the drawing of fiat outer boundaries in the spatial realm which yields fiat objects. A bona fide object, instead, has its spatial delimitations defined *per se*, possibly

<sup>4</sup>The term 'fiat' comes from Latin, and is defined by Merriam-Webster as "a command or act of will that creates something without or as if without further effort".

existing even before the presence of cognitive subjects in Earth. Examples of genuine objects are: you and me, a tree, the planet Earth. Examples of fiat objects are: upper, middle and lower femur, all non-naturally demarcated geographical entities, including Colorado, the North Sea and so on (Ibid., p. 6).

According to all of this, we understand that a **Body surface region** is a subtype of *fiat boundary*. This entity is actually hard to classify exclusively under one of the two rubrics (*bona fide* or *fiat*). This is because, on the one hand, if every region (part) of the body surface is still a *bona fide* boundary in a sense, as it is genuinely separated from the outer environment, on the other hand only a cognitive act (we are considering that the region is still tied to the body) can delineate such region. We adopt here a convention in which if there is any fiat demarcation, then the object under analysis is a fiat object.

In UFO, parasitic substantials such as stains, edges, bumps - which are named *features* in DOLCE (49), as well as what Pribbenow terms a *negative object* (e.g., a hole, the interior of a drawer) (125) are not considered *moments*, but *objects*. Hence, the relation between those entities and their hosts is one of *inseparable parthood*, not one of *inherence* (34, p. 216). For instance, we say that a hole in a piece of cheese is an inseparable part of the cheese, as opposed to one of its moments. That is, the hole is not exactly existentially dependent from the cheese, but a part of it with the same lifecycle. That being said, we submit that the body surface is an inseparable part of the body, in contrast to the relation of *bounds* used by the FMA to state that the body surface bounds the body.

The category of **Material anatomical entity** is partitioned here as shown in Figure 16. By looking at that model, we quickly reach the *kinds* **Human body**, **Heart** (type of **Organ**) and **Cardiovascular system** (type of **Organ system**). These entities are modeled as such in virtue of the meta-properties they hold, namely, those of rigidity, unity, persistence and the provisioning of an identity principle. We then start to elaborate on the *category* of **Organ component** as depicted in Figure 17.

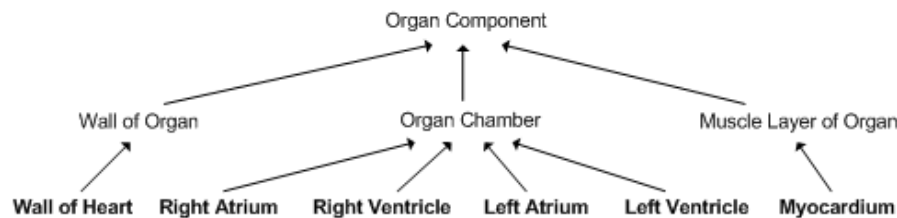


Figure 17: The *is-a* taxonomy descending from the *category* **Organ component**. The *kind* universals are those highlighted in boldface.

The **Wall of heart** is a three-layered object that is a type of **Wall of organ**. Its three layers (which are parts of it) are the endocardium ('endo' = within + 'heart'), epicardium ('epi' = upon + 'heart') and *myocardium* ('myo' = muscle + 'heart') (123, p. 55). The wall of the heart is continuous with the walls of the systemic and pulmonary arterial and venous trees (94). The **Myocardium** is a **Muscle layer of organ** that is capable of holding contractions. Unlike all other types of muscle cells (myocytes), cardiac muscle cells: (i) branch, (ii) join at complex junctions called intercalated discs so that they form cellular networks, and (iii) each contain single, centrally located nuclei. It is still worth to highlight that though sometimes used in an ambiguous way, a cardiac muscle cell is not a fiber. The term cardiac muscle fiber, when used, refers to a long row of joined cardiac muscle cells (123, p. 56). Figure 18 illustrates in detail the wall of the heart by emphasizing the different sorts of tissue that compose each of its three layers.

One of the most important aspects of the heart is that it is a chambered organ. The heart has four chambers

(instances of *kinds*), viz., the right and left atria, and the right and left ventricles, see Figure 17. This chambered aspect of the heart is relevant here, firstly, to provide even a picture of this important point of view of the heart, and secondly, for supporting a better understanding of the myocardial subdivisions introduced in the sequel that hold different characteristics in what concerns the atrial and ventricular parts. In Section 5.3, the cardiac circulation is also put in perspective by relying on the chambered structure of the heart.

We now shift our analysis to a fiat *category*, viz., **Region of organ component**. A **Region of wall of heart** can be delimited by an arbitrary demarcation according to the structures which define its cavities. Among them, those which are of our concern are the four chambers of the heart. Accordingly, the walls of the right and left atria and those of the right and left ventricles. As the **Myocardium** is part of the **Wall of heart** (we present a parthood taxonomy in the following), the fiat subdivisions of the latter have as parts, respectively, those fiat subdivisions of the former. Namely, right and left atrial myocardium, and right and left ventricular myocardium. All of them are universals of the type *kind*, in conformance with Figure 19 below.

We can then finally reach more specific objects in the heart anatomy which are directly active in the electrophysiological phenomena of our interest. We then need to elaborate the *category* **Portion of tissue** as illustrated in Figure 20. Among several sorts of tissue existing in the human body, the muscle tissue is of our interest here, in contrast to epithelium, connective tissue, neural tissue, or even heterogeneous tissue. The muscle tissue, however, may be either smooth or striated. Moreover, the latter may be skeletal (attached to bones), or a special one, namely, the *cardiac muscle tissue*. Although this navigation through the typology of portion of tissue might be enlightening in this text, we shall not include the subtypes just mentioned in our theory to avoid unnecessary prolixity. What is worth enough not only for mentioning, but also for formal treatment here is a closer look at what a portion of tissue is made of. As put by Geselowitz, “[b]ody tissues are made of cells immersed in a fluid matrix” (22, p. 859). We then submit, for further application in this text, that any type of **Portion of tissue** is *constituted* by its cells in addition to its extracellular matrix. The notion of *constitution* is elaborated still in this section.

Figure 20 shows three subtypes of portion of tissue. All of them have spatial boundaries defined by means

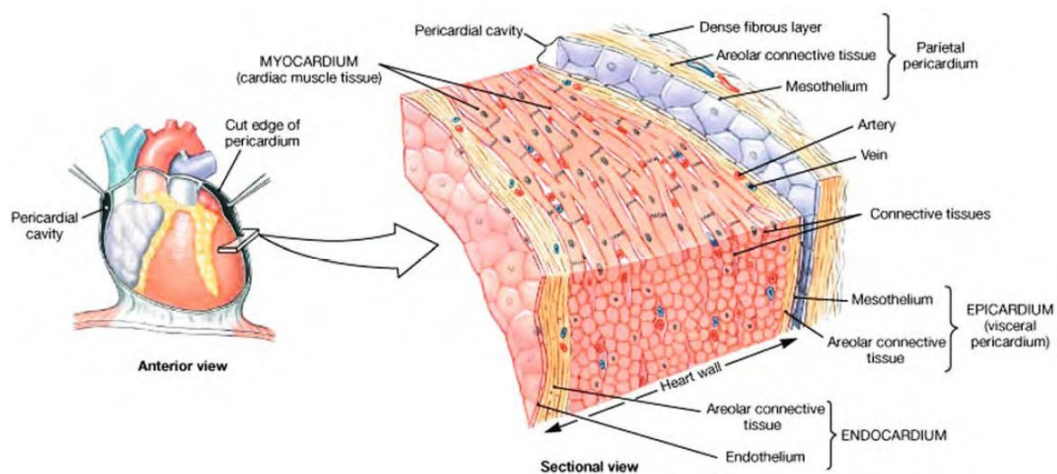


Figure 18: Internal anatomy of the wall of the heart. It contains three layers: the superficial epicardium; the middle myocardium, which is a muscle layer; and the inner endocardium. Note that cardiac muscle cells contain intercalated disks that enable the cells to communicate and allow direct transmission of electrical impulses from one cell to another. Source: (123, p. 58) borrowed from Human Anatomy, 4th Ed. by Frederic H. Martini, Michael J. Timmons, and Robert B. Tallitsch.

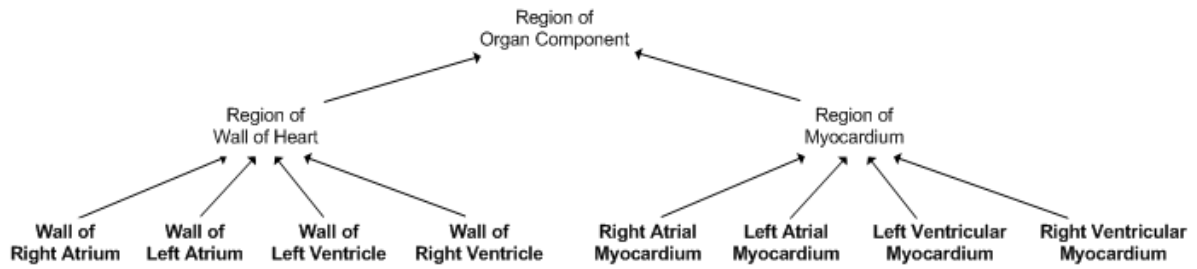


Figure 19: The *is-a* taxonomy descending from Region of organ component. Again, the substantial sortals (*kinds*) are set in boldface.

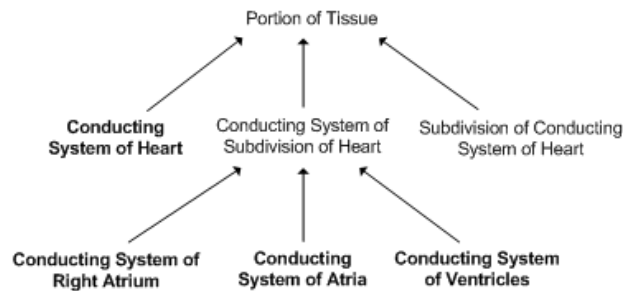


Figure 20: The *is-a* taxonomy descending from Portion of tissue. This picture follows the same convention w.r.t. the exhibition of *kinds* in boldface in contrast to *categories*. The *category* Subdivision of conducting system of heart is further elaborated.

of a fiat demarcation. The *kind* Conduction system of heart comprises the whole portion of cardiac muscle tissue which bears properties that enable it to conduct an electrical impulse. That is, this portion of tissue has special conducting properties that characterize it as such, cf. Figure 18. It can, however, be further distinguished by several subdivisions of the conducting system of the heart which are elaborated in Figure 21 that follows. In addition, the *category* Conduction system of subdivision of heart denotes a fiat medical division like the Conduction system of ventricles - CSV henceforth, the Conduction system of atria - CSA henceforth, or even more specifically, the Conduction system of right atrium. We shall see in Section 5.3 the relevance of such fiat divisions as they gather interesting subsets of functionality.

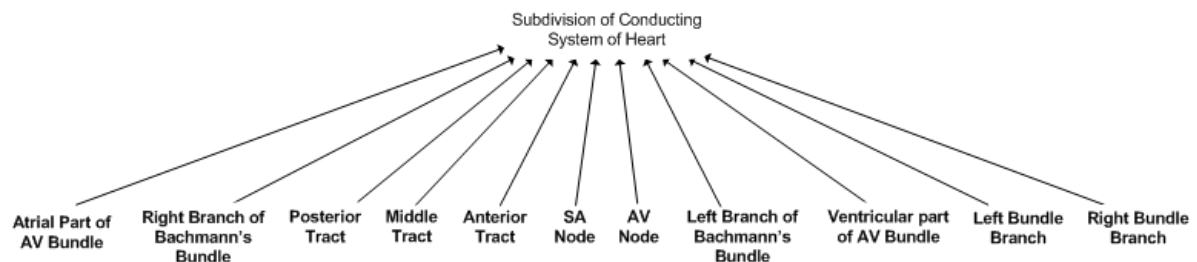


Figure 21: Referred subdivisions of the conducting system of the heart. These *kind* entities are indeed ultimate parts of the conducting system of the heart (at a mesoscopic level of granularity) as described further on in this text.

We refer also to Figure 22 in order to provide an anatomical picture of the conducting system of the heart. Thereby, it may be easier to apprehend the several subdivisions (mostly fiat) that compose it. The sinoatrial node, or just **SA node**, is located in the so-called “roof” of the right atrium. It has short dimensions compared to other regions of the heart conducting system (it indeed cannot be seen grossly); but to anticipate Section 5.3, it is of

foremost importance in the system.

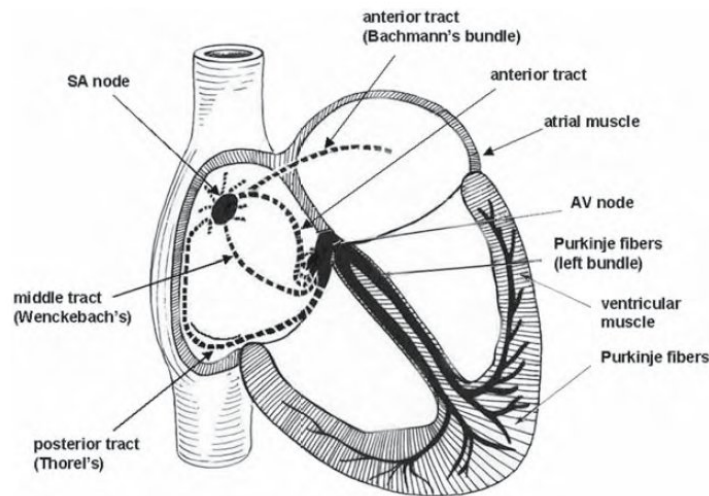


Figure 22: The conducting system of the heart - source: (126, p. 124). Normal excitation originates in the sinoatrial (SA) node, then propagates through both atria (internodal tracts shown as dashed lines). The atrial depolarization spreads to the atrioventricular (AV) node, passes through the bundle of His (not labeled), and then to the Purkinje fibers (cf. Section 5.3), which make up the left and right bundle branches; subsequently, all ventricular muscle becomes activated.

Three preferential anatomical conduction pathways have been reported from the SA node to the atrioventricular node (AV node). Namely, the **Anterior tract**, **Middle tract** (or Wenckebach pathway) and **Posterior tract** (or Thorel pathway), see Figure 22. These are the shortest electrical routes between the nodes. They are microscopically identifiable structures, appearing to be preferentially oriented fibers, that provide a direct node-to-node pathway. The anterior tract, in particular, extends from the anterior part of the SA node, bifurcating into the so-called Bachmann's bundle and a tract that descends along the right atrium and connects to the AV node, see Figure 22. Nonetheless, the terms assigned to these entities are controversial. Both of the three are referred to as *anterior tract*. By considering all of our bibliographic references, we have adopted the following. The entity resulting from the bifurcation aforementioned which descends to the AV node seems not to be referred to apart from that initial part tied to the SA node. Therefore, we only consider here two entities. One of them is the entity topologically connected to the SA node that continually descends till reaching the AV node. This is the **Anterior tract**. We then convey as **Bachmann's bundle** the entity topologically connected to the Anterior tract (at that bifurcation point) that spans from the left to the right atrial myocardium. The Bachmann's bundle is indeed the only electrical route to the left atrial myocardium.

Since we are talking about *conductor* anatomical entities, it is worth to anticipate that the conduction velocity varies considerably in the heart and is directly dependent on the diameter of the myocytes. For example, current conduction is greatly slowed as it passes through the AV node. This is mainly because of the small diameter of its nodal cells and the tortuosity of the cellular pathway they form. However, it seems like everything in the (canonical) heart anatomy and physiology is on the right place at the right time, and it is not different with this delay. It is actually strategic to allow adequate time for ventricular filling, cf. Section 5.3. The AV node is located in the so-called "floor" of the right atrium.

The next subdivision of the heart conducting system of our concern is the Bundle of His, or AV bundle.



We have abide by the latter term to comply with the FMA<sup>5</sup>. The AV bundle is bifurcated into two other entities named **Right bundle branch** and **Left bundle branch**, see Figure 22. The complex network of conducting fibers that extends from either the right or the left bundle branches is composed of rapid conduction cells that emerge as the so-called *Purkinje fibers*. The Purkinje fibers in both the right and the left ventricles act as preferential conduction pathways to provide rapid electrical conduction within the various regions of the ventricular myocardium. “Purkinje fibers”, however, is actually a term used to distinguish specific portions of tissue (viz., the right and left bundle branches) that bear the disposition of conducting an electrical impulse very fast. In Section 5.3, such functional aspects of the heart conducting system are discussed in detail.

The two material anatomical entities left (viz., **Anatomical cluster**, **Portion of body substance**) are introduced in what follows. Hitherto, we have presented anatomical entities relevant in the ECG domain. They have been introduced in the context of subsumption taxonomies in an attempt to favor comprehensibility of their nature. Nevertheless, perhaps the most characteristic aspect of those anatomical entities is that they form a comprehensive (stable) mereological, even mereotopological whole. Ergo, we now focus on the parthood relationships held among the anatomical entities introduced above. Our mereological analysis starts by looking *at* the **Human body** (the universal), and then elaborates on the parts that compose the human **Heart** - see Figure 23 for our anatomical parthood taxonomy.

In the anatomical partonomy of Figure 23, we use the parthood relation by adopting what is known, in Formal Ontology, as *ground mereology*, cf. (34, Chapter 5). The *part of* links shown in Figure 23 represent a class-level relation (holding between two classes, e.g. the **Right atrium** is part of the **Heart**) defined from an instance-level *part of* (holding between two individuals, e.g., my right atrium is part of my particular heart). The class-level parthood is defined by accounting for the instance-level version. The latter is a primitive relation characterized by the meta-properties of *irreflexivity*, *asymmetry* and *transitivity*. Formally, this means that:

**Irreflexivity:**  $\forall c, t \quad \neg \text{part\_of}(c, c, t)$

**Asymmetry:**  $\forall c_1, c_2, t \quad \text{part\_of}(c_1, c_2, t) \rightarrow \neg \text{part\_of}(c_2, c_1, t)$

**Transitivity:**  $\forall c_1, c_2, c_3, t \quad \text{part\_of}(c_1, c_2, t) \wedge \text{part\_of}(c_2, c_3, t) \rightarrow \text{part\_of}(c_1, c_3, t)$

The class-level parthood (the links in Figure 23) can then be obtained as follows.

$\text{part\_of}(C_1, C_2) =_{\text{def}} \forall c_1, t ( \text{instance\_of}(c_1, C_1, t) \rightarrow \exists c_2 ( \text{instance\_of}(c_2, C_2, t) \wedge \text{part\_of}(c_1, c_2, t) ) )$

The formula above defines a permanent parthood relation, in the sense that, first, if it is the case that every instance of  $C_1$  exists at some time instant  $t_1$  ( $C_1$  is not abstract), then whenever they exist, exist as part of some  $c_2$ .

The inverse relation *has part* (or *reciprocal*, if we follow GALEN’s (127) convention, holding for all *part of* assertions above) can be defined by the same token by only inverting the domain and image of its *part of* counterpart. Notice in Figure 23 that some entities in the partonomy have only one part. Although this is not a problem when adopting ground mereology, the real reason here is something else. Those entities do have other universals as parts, but these are not relevant for the representation of the ECG. Finally, it is important to highlight that we have used the instance-level relation of *part of* here to represent a *proper parthood* relation. If necessary, an instance-level *improper parthood* relation<sup>6</sup> can be defined as usual:

<sup>5</sup>Actually, there is evidence enough collected by Laske and Iaizzo in (126, p. 131) for further distinguishing the AV bundle and Bundle of His as two different entities. However, this is only meaningful in a strongly detailed modeling that takes account of lower levels of granularity. In face of this, we have found preferable to adhere to the FMA modeling.

<sup>6</sup>This is actually equivalent to the most general version of *part of* which is reflexive, anti-symmetric and transitive.

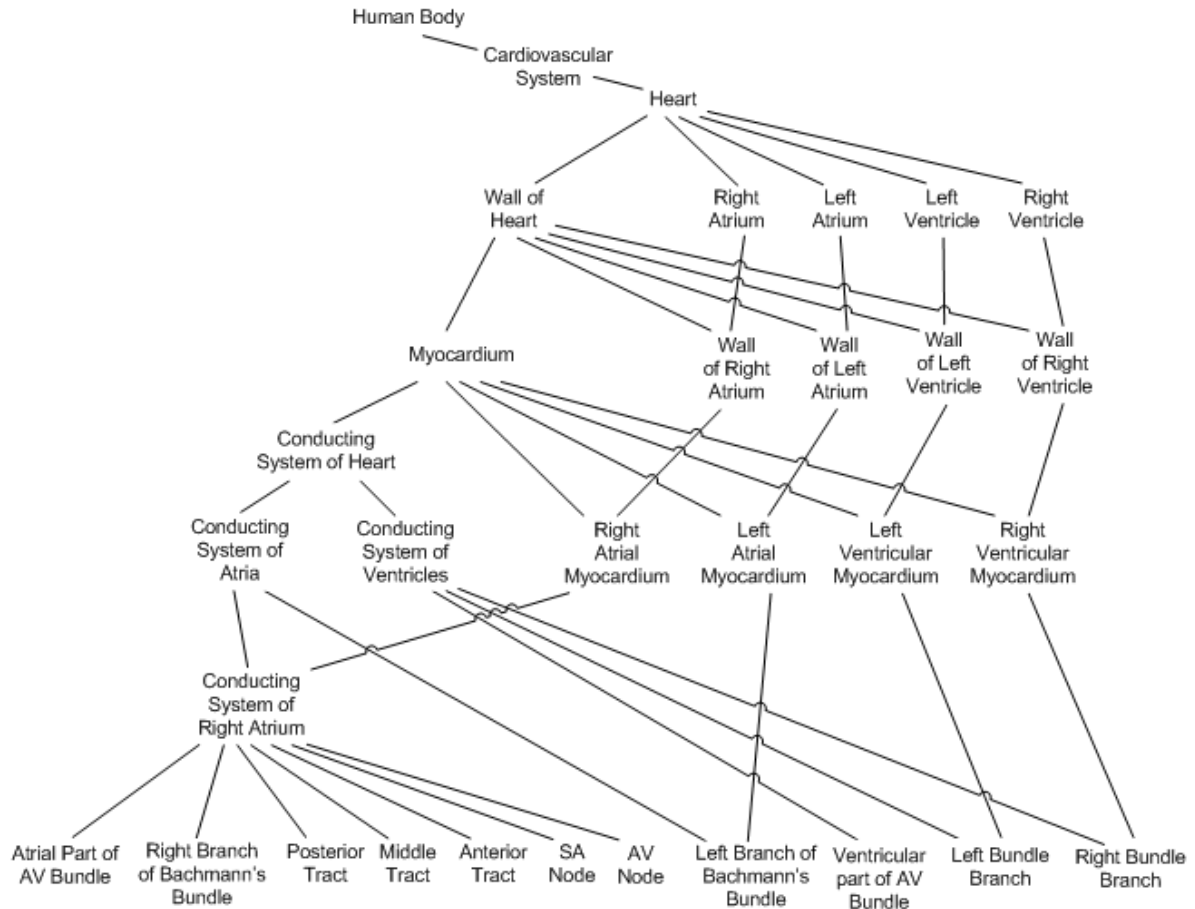


Figure 23: Partonomy of anatomical entities which concern the ECG, inspired in the FMA. The lines represent (proper) *part of* relationships (from the bottom to the top) between the anatomical entities - all of them are substantial sortals (*kinds*). Its inverse relation *has part* also holds from the top to the bottom for all lines. The heart has as parts the right and left atria, the right and left ventricles, and the **Wall of heart**. As said before, the latter in turn has as parts the layers of endocardium, epicardium and **Myocardium**. This type of **Muscle layer of organ** is further divided (not completely) in right and left atrial myocardium, and right and left ventricular myocardium. These anatomical entities finally have as parts the conducting systems of right and left atria, and right and left ventricles, respectively. We explicitly include here only the **Conducting system of right atrium**, since it exemplifies a full division into multiple ultimate parts of the heart conducting system in our scope.

$$\mathbf{improper\_part\_of}(c_1, c_2, t) =_{def} \mathbf{part\_of}(c_1, c_2, t) \vee (c_1 = c_2 \text{ at } t)$$

Unlike the FMA curators, we have not included in one single anatomical partonomy entities at *different levels of granularity* - cf. (95), e.g., a single *SA node myocyte*<sup>7</sup>. We submit that such a universal is a *grain of* the collective of *SA node myocytes*, not a *part of* the *SA node*; and *grain of*<sup>8</sup> is not the usual *part of* (95, 34). The collective of *SA node myocytes* in turn *constitutes* the *SA node*, in addition to the extracellular fluid matrix of the *SA node*. The latter then emerges from the mereological sum<sup>9</sup> of its cells and its extracellular matrix (ECM), just as any other *Portion of tissue*. However, we need the collective of *SA node myocytes* to be still partitioned by means of the relation *subcollection of* into two sub-collectives, viz. *Pacemaker SA node myocytes* and *Transitional*

<sup>7</sup>In fact, such an entity is not considered in our ECG theory as far as our scope do not cover the cellular level. For illustrating the choice made by the FMA curators, the entity *Pacemaker cell of sinuatrial node* (FMAID:83383) is a *part of* the *Sinuatrial node* (FMAID:9477) (94, access on March 12, 2009).

<sup>8</sup>We make use of the notion of *grain of* in the section dealing with the ECG. In that section it is provided a formal description of this relation.

<sup>9</sup>The sum  $z$  of two objects  $x$  and  $y$ , symbolized as  $Sum(z, x, y)$ , is the entity such that every object that overlaps with  $z$ , overlaps either with  $x$  or with  $y$  (or with both) (34, p. 147). The notion of overlaps is defined further on in this text.

SA node myocytes<sup>10</sup>. This partition is made according to their specific properties further discussed in Section 5.3.

Figure 24 shows the subsumption hierarchy for the two anatomical categories not presented yet, those of Anatomical cluster and Portion of body substance. They cover the entities Cell cluster and Portion of extracellular matrix, i.e., a fluid matrix a Portion of tissue is made of (or *constituted of*). In face of the aforementioned ontological distinctions, however, we start to make use of OntoUML in an effort to make them explicit. Figure 25 depicts, on the top, the relation *subcollection of* as it holds between the SA node myocytes and the Pacemaker SA node myocytes, on one side, and between the former and the Transitional SA node myocytes, on the other. It is also shown in Figure 25 the CSA Myocytes and CSV Myocytes, in virtue of their relevance for the analysis carried out in the next section.

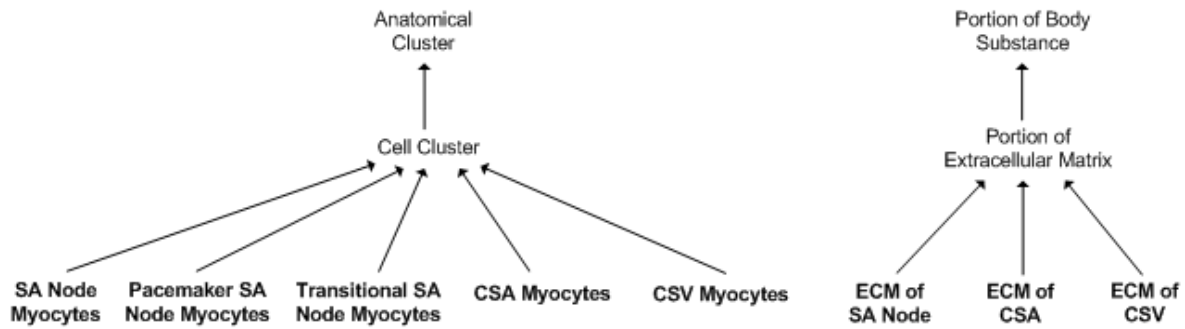


Figure 24: Material anatomical *categories* Anatomical cluster and Portion of body substance. All of the *collectives* of myocytes (in boldface) are subtypes of the *category* Cell cluster. On the other hand, the *quantities* constituting different portions of tissue in the heart (also in boldface) are subtypes of the *category* Portion of extracellular matrix (ECM stands for “extracellular matrix”).

The essential parthood exemplified in this model is defined further on in this text. The notions of *collective* and *subcollection of* are contemplated in the UFO ontology (34, Section 5.5), and the former is also discussed in depth by Rector et al. in (95). The *subcollection of* relation is a specific type of parthood relation - cf. (34, Section 5.5), and is also characterized by the meta-properties of *irreflexivity*, *asymmetry* and *transitivity* as defined above, but additionally by *weak supplementation*.

$$\text{Weak supplementation: } \forall c_1, c_2, t ( \text{subcollective\_of}(c_1, c_2, t) \rightarrow \exists c_3 ( \text{subcollective\_of}(c_3, c_2, t) \wedge \neg \text{overlaps}(c_1, c_3, t) ) )$$

Mereological overlapping can be in turn defined as usual.

$$\text{overlaps}(c_1, c_2, t) =_{def} \exists c_3 ( \text{part\_of}(c_3, c_1, t) \wedge \text{part\_of}(c_3, c_2, t) )$$

The notion of *constitution* is formally described by Masolo et al. in (49, p. 32 and 34). However, their axiomatization makes use of many other formulae and notions that happen to be better comprehended in the context of the DOLCE ontology as a whole. For us, it will suffice to assert, first of all, that constitution is not identity. A classical example to illustrate this can be given by a portion of clay that *constitutes* a statue (made) of clay. Since the statue can undergo replacements of certain parts, but an amount (portion) of matter can not, they can not be the same (49, p. 21). Constitution, actually, stands for the meta-properties of *irreflexivity*, *asymmetry* and *transitivity*. We can then build upon the relation *constituted by* to characterize a more specific type of constitution, viz., the relation *partially constituted by*. This means that (it is explicitly known that) there is at least two entities a

<sup>10</sup>The collective of SA node myocytes is divided roughly in a half-and-half proportion into pacemaker and transitional cells (126, p. 124).

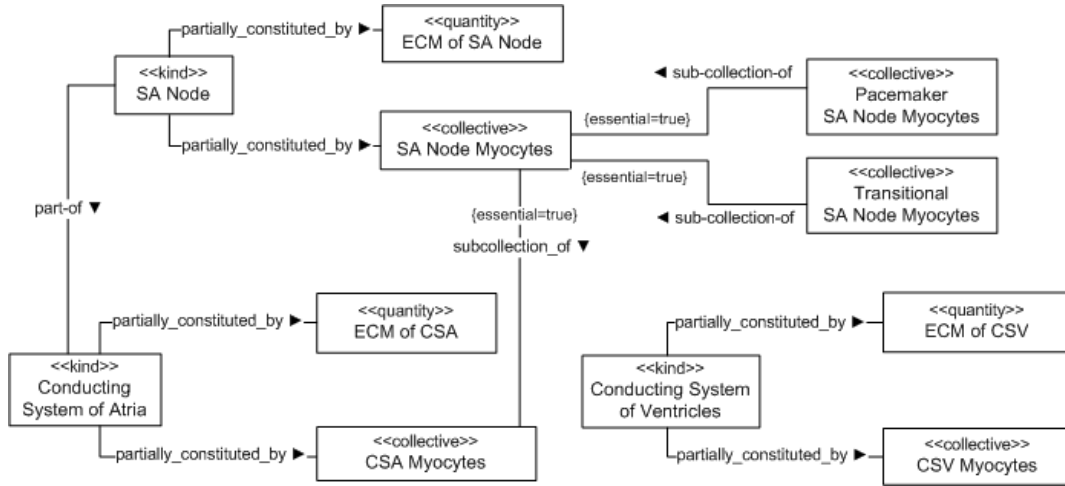


Figure 25: Relations involving myocytes of subdivisions of the heart conducting system. The SA node myocytes are further distinguished into two collectives in virtue of specific properties held by pacemaker myocytes on one side, and transitional myocytes on the other. Both the *part of* and *subcollection of* assertions have their inverse relations *has part* and *has subcollection* holding as well. The subcollections of the SA node myocytes are actually *essential parts* of this collective. Note that: (i) the SA node is actually a direct *part of* the Conducting system of right atrium, which is *part of* the Conducting system of atria; the SA node is then, by transitivity, also *part of* latter as asserted in this model; (ii) by the same token (i.e., transitivity propagation through the myocytes of the conducting system of right atrium), the SA node myocytes happens to be an essential *subcollection of* the CSA Myocytes.

given third entity is constituted by. This relation holds the meta-properties just mentioned as well and is formally characterized as follows at the instance-level.

$$\text{partially\_constituted\_by}(c, c_1, t) =_{\text{def}} \text{constituted\_by}(c, c_1) \wedge \exists c_2 (\neg \text{overlaps}(c_2, c_1) \wedge \text{constituted\_by}(c, c_2))$$

Notice that the overlapping relation used here does not rule out the possibility of two or more constituents which are mixed together. For example, wine is partially constituted by alcohol and grape, even though these two substances can not be distinguished by the naked eye when mixed together. They, however, are still two different substances that do not overlap under a chemical perspective. At the class-level, the relation *partially constituted by* turns to be:

$$\text{partially\_constituted\_by}(C_1, C_2) =_{\text{def}} \forall c_1, t (\text{instance\_of}(c_1, C_1, t) \rightarrow \exists c_2 (\text{instance\_of}(c_2, C_2, t) \wedge \text{partially\_constituted\_by}(c_1, c_2, t)))$$

Besides, one might notice that the *Body surface* and *Body surface region* entities have not been included in the parthood of Figure 23. The reason is that the mereological relation between the former and the *Human body* is not a standard (class-level) *part of*, but that specific (stronger) one termed *inseparable part of* by Guizzardi in (34, p. 216) as depicted in Figure 26. Likewise, we assert the essential parthood to relate (say) the SA node myocytes to the CSA myocytes by means of the *subcollection of* mereological relation. Both of these relations have been formally defined in UFO by using an intensional quantified modal logics as mentioned in Section 4.1.1. We define them here in FOL (with some loss in expressiveness) by taking account of time as follows. Firstly, we refer to the instance-level.

$$\text{inseparable\_part\_of}(c_1, c_2) =_{\text{def}} \forall t (\text{exists}(c_1, t) \rightarrow \text{part\_of}(c_1, c_2, t))$$

$$\text{essential\_part\_of}(c_1, c_2) =_{def} \forall t ( \text{exists}(c_2, t) \rightarrow \text{part\_of}(c_1, c_2, t) )$$

At the class-level, those relations turn to be defined as follows.

$$\text{inseparable\_part\_of}(C_1, C_2) =_{def}$$

$$\forall c_1, t ( \text{instance\_of}(c_1, C_1, t) \rightarrow \exists c_2 ( \text{instance\_of}(c_2, C_2, t) \wedge \text{inseparable\_part\_of}(c_1, c_2, t) ) )$$

$$\text{essential\_part\_of}(C_1, C_2) =_{def}$$

$$\forall c_2, t ( \text{instance\_of}(c_2, C_2, t) \rightarrow \exists c_1 ( \text{instance\_of}(c_1, C_1, t) \wedge \text{essential\_part\_of}(c_1, c_2, t) ) )$$

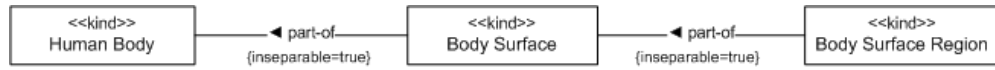


Figure 26: Relations involving the Body surface. We make use of the *inseparable parthood* distinction to characterize the parthood holding between Human body and the Body surface, and then between the latter and a Body surface region. Unlike the Human body and Body surface, the inverse *has part* relation does not hold between the Body surface and the Body surface region.

Altogether, each *part of / has part* relation presented above has been carefully asserted. Parthood propagation over our anatomy model is then safely warranted. This anatomy model for the ECG consists basically of the *parthood* and *is-a* taxonomies presented above. A complete documentation of it (e.g. containing class ID and textual definitions) is given in Section 5.7. It may be worthwhile to recall, however, that the anatomy model conveyed here is not intended to be used aside from the ECG Ontology. In other words, we do not regard it as an ontology of cardiac anatomy *per se*.

## 5.3 Heart Electrophysiology

Once we have established the required anatomical (structural) basis, we can concentrate our inquiry on what directly concerns the ECG, viz., the heart electrophysiology. The contents of this section are based on Laske and Iaizzo's chapter dealing with the heart conduction system (126), on Guyton and Hall's physiology textbook (128), as well as on a remarkable article written by Geselowitz (22) that presents the theory of the ECG from a mathematical perspective.

A usual explanation of the heart electrophysiology - as encountered in the references just mentioned - starts with an introduction of the phenomena that take place in lower levels of physiological activity. Basically, bioelectric sources spontaneously arise in the heart at the cellular level. The heart myocytes (muscle cells) are immersed in an extracellular fluid separated from their interior by their membranes, which carry out a control of ions transport. In the resting state, the interior of the myocytes has a negative potential with respect to the exterior, i.e., these cells are electrically polarized.

However, particularly in the sinoatrial (SA) and atrioventricular (AV) nodes - parts of the conducting system of the heart, see Figure 23, special myocytes bear the disposition for abruptly depolarizing, and then returning back to their resting value. This phenomenon is called "self-excitation", since these particular cells spontaneously open their membranes for exchanging ions with the extracellular fluid matrix. The special myocytes we are talking about have been named *pacemaker cells*; according to the current medical knowledge, they do not exist in any other region of the human body except in the heart (22, 128). As a result of their self-excitation that lead to such depolarization, the SA and AV node pacemaker myocytes give rise to action potentials (or electrical impulses)

which are propagated to its neighboring myocytes over the myocardial conducting tissues and normally reach the entire heart, see Figure 27. That is why the SA and AV nodes are called the heart pacemakers. However, since this sort of electrical impulse arises in the SA node at a faster rate<sup>11</sup>, the AV node electrical impulse is said to be overdriven by the SA node impulse (22). Thus, the Cardiac electrical impulse is referred to as being the electrical current generated by the SA node pacemaker myocytes that is conducted throughout the Conducting system of heart.

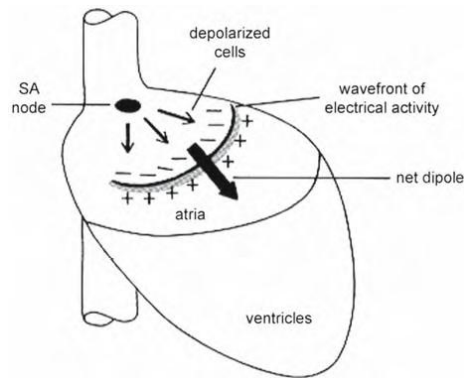


Figure 27: Propagation of the cardiac electrical impulse. After conduction begins at the sinoatrial (SA) node, cells in the atria begin to depolarize. This creates an electrical wavefront that moves down toward the ventricles, with polarized cells at the front, followed by depolarized cells behind. The latter stimulate the former to become depolarized as well. The separation of charge results in a dipole across the heart (the large black arrow shows its direction). Source: (129, p. 192), modified from D.E. Mohrman and L.J. Heller (eds.), *Cardiovascular Physiology*, 5th Ed., 2003.

For conveying the cardiac electrical impulse (CEI, henceforth) around the heart, there are myocytes in addition to the SA and AV pacemaker myocytes that compose such conducting system of the heart. They are named *transitional myocytes* (see Figure 25), due to their capability for propagating the CEI to their neighbors. Ergo, if on one side pacemaker cells have an excitatory nature, transitional cells have on the other side a conductive one. Either pacemaker or transitional myocytes go through two different electrical states as clarified in the caption of Figure 27. These are the polarized and depolarized states, which qualify all subsets of myocytes in the conducting system of the heart, see Figure 28). All the myocytes are polarized in their resting state. Pacemaker myocytes spontaneously depolarize and then stimulate their neighbors to depolarize as well. The transitional myocytes, however, are those whose main feature is to conduct the action potentials that constitute the CEI. Overall, as the CEI has been conducted over them, both pacemaker and transitional cells *repolarize*, i.e., they return to their resting polarized state.

The major conducting pathway of the heart is the so-called His-Purkinje system, see Figure 22. It is composed by the atrioventricular bundle (AV bundle, or bundle of His), then bifurcated into the left and right bundle branches - constituted by the Purkinje fibers. As a response to the CEI conducted over that system, the myocardium holds contractions in its atrial and ventricular parts for pushing blood respectively into the ventricles and either into the systemic or pulmonary circulation. Although the circulatory phenomenon fueled by the heart is central in the sense that it is actually what emerges from the organ functioning in lower levels of activity, we shall not elaborate on it here since we are in fact interested in the ECG underlying phenomena, viz., those of electrophysiological nature

<sup>11</sup>While the pacemaker rate of the SA node varies between 60-100 bpm (beats/min), that of the AV node stays usually between 40-55 bpm (126, p. 127). If there is some problem with the impulse generated by the SA node - say, it has been *blocked* in its atrial pathway to the AV node, then the one generated by the AV node succeeds to give rise to a *escape beat*. This beat can actually temporally keep one alive till the *heart block* in the atrial pathway is fixed.

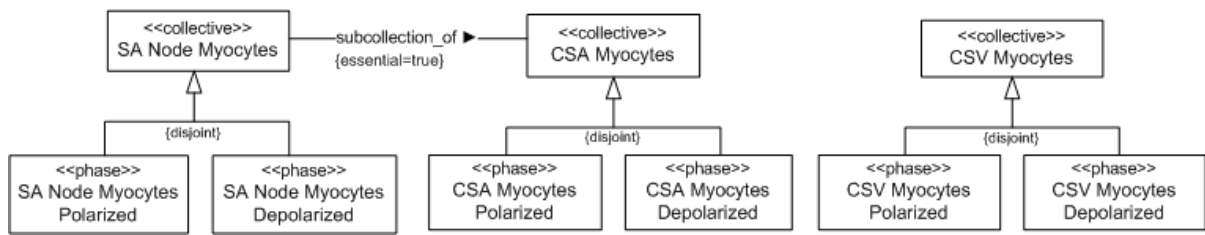


Figure 28: Polarized and depolarized phases which every *collective* of myocytes in the conducting system of the heart alternate between. Another reasonable phase might be considered as being the phase in which the cells are not entirely polarized neither depolarized, but moving from one state to the other. A *phase* universal stands for the meta-properties of *anti-rigidity* and *relational independence*, and it carries a principle of identity inherited from a unique a kind universal, collective or quantity. In this case either of the two phases inherits the principle of identity of the collective of SA node myocytes.

which are directly mapped in an ECG waveform. Nevertheless, Figure 29 provides a general account of the cardiac circulation important for understanding the heart functioning as a whole.

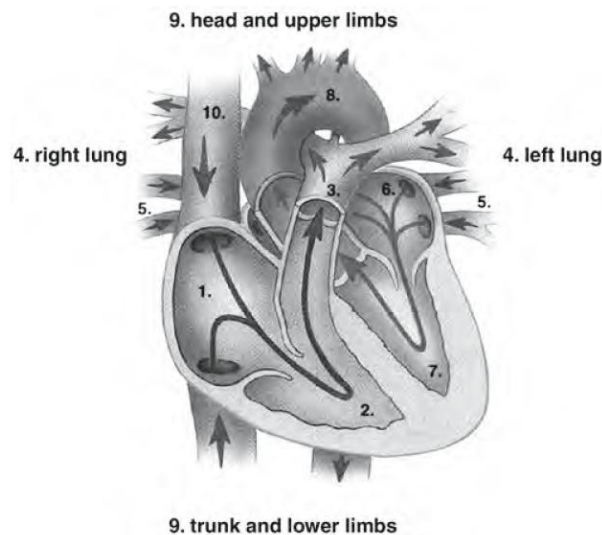


Figure 29: Cardiac circulation. This picture is not required for the reading that comes in the rest of this thesis, but finds place here for the sake of a broader comprehension of the heart functioning. The numbers in the picture guide the reading that follows. Blood collected in the right atrium is pumped into the right ventricle. On contraction of the right ventricle, blood passes through the pulmonary trunk and arteries to the lungs. The left atrium pumps the blood into the left ventricle. Contraction of the left ventricle sends the blood through the aortic artery to all tissues in the body. The release of oxygen in exchange for carbon dioxide occurs through capillaries in the tissues. Return of oxygen-poor blood is through the superior and inferior venae cavae, which empty into the right atrium. Note that a unidirectional flow of blood through the heart is accomplished by valves. Source: (123), in a reprint from “Principals of Human Anatomy”, by G.J. Tortora, 1999 Biological Sciences Textbooks.

Given that brief overview, we now focus on an ontological representation of the human heart electrophysiology meaningful for the representation of the ECG. For this, we build upon the Ontology of Functions (OF) proposed by Burek et al. (4) as introduced in Section 4.1.3. Basically, we aim at providing a clear structure of heart electrophysiological functions (what they are), and how and by whom can they be realized. We also intend, by these means, to be able to further reconstruct those physiological entities from a particular ECG record.

In conformity to the textual description given above, we have selected some functions which are more directly related to the ECG data. In other words, the functions the ECG provides direct mapping information for. These

are To generate CEI, To conduct CEI and To restore EPs - EPs stand for electrical potentials. Figures 30, 31 and 32 that follows provide, respectively, their representation in the OF schema.

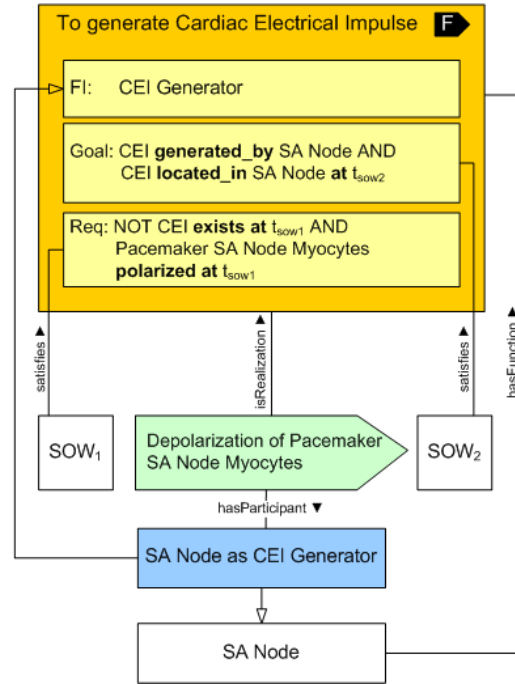


Figure 30: Function To generate CEI represented in the OF framework. This function is *realized by* means of the process of Depolarization of pacemaker SA node myocytes. This process is started if the Pacemaker SA node myocytes are polarized and there is no CEI existing in the whole heart conducting system at  $SOW_1$ . In a normal scenario, the process is finished at  $SOW_2$  satisfying the goal that the CEI has been *generated by* the SA node as CEI generator and it is (still) *located in* the SA node at this point. Note that location entails existence.

Those functions are put together in the model of heart electrophysiological functions, see Figure 33. Their applicability is further clarified in Section 5.6. Note that the notion of function falls into the UFO ontology under the rubric of *mode*. We then use the *characterized by* relation set from a *kind* towards a function in place of OF's *has function*. Our rationale is to keep as much as possible everything understood in terms of the UFO signature. The notions of *mode* as well as *characterization* are elaborated on in what follows. Besides, the function To conduct CEI is taken in account here as it is manifested in the *atria* and *ventricles*. This could be refined to cover the CEI conduction over the specific subdivisions of the heart conducting system introduced in the previous section - e.g., internodal tracts, bundle branches, etc. Nonetheless, as we shall see further on, the ECG provides a direct picture of the CEI conduction around the atria and ventricles. Probabilistic mapping inferences could be made, say, for predictions about the CEI conduction throughout the AV node by taking account of indirect associations between different elements of the ECG waveform. This goes beyond the scope of this thesis, but is discussed in terms of future work (cf. Chapter 9).

The function To generate CEI (on the top in Figure 33) *characterizes* the SA node - it is one of its properties, and is *realized by* the process of Depolarization of pacemaker SA node myocytes. The SA node as CEI generator (a type of CEI generator) *participates in* that process, in addition to the CEI itself. However, while the former participates in the process for its entirety, the latter participates at and only at its last instant (cf. definition of *generated by* given below). The class-level *has participant* relation as used in this model is defined below as well. The function To conduct CEI (on the bottom) in turn *characterizes* both the conducting systems of atria and ventricles, but also the SA node. This function can be *realized by* both the process of Depolarization of



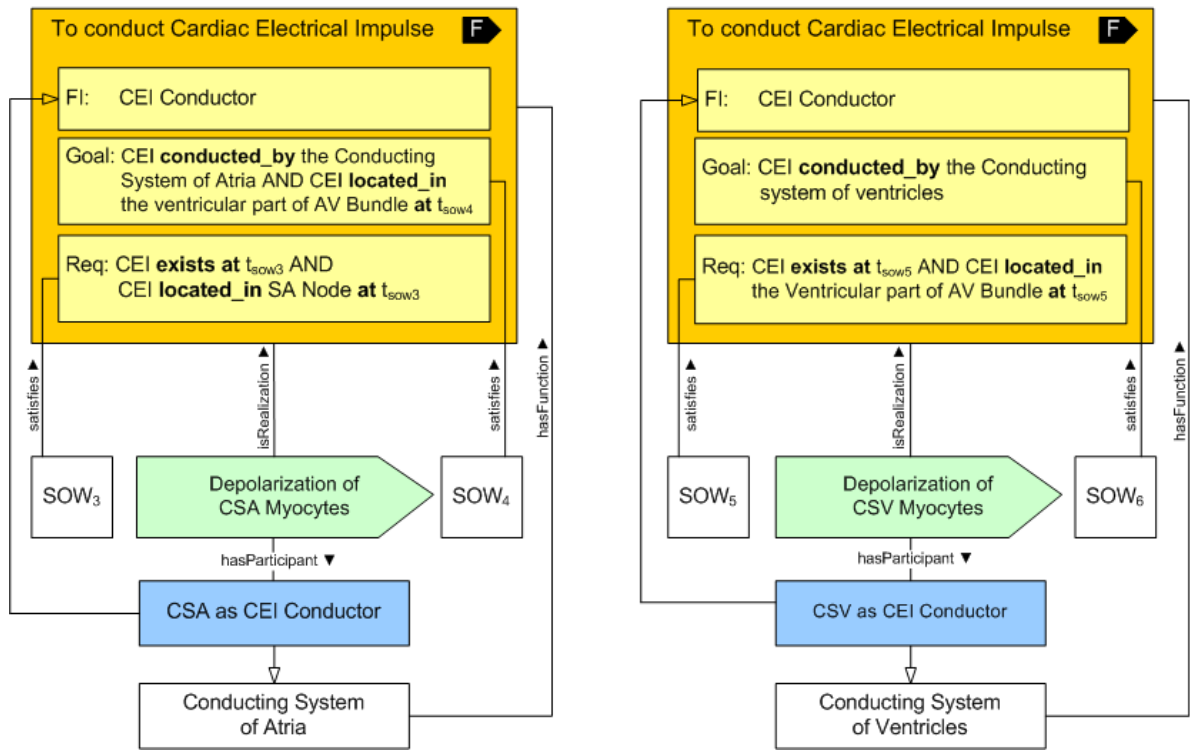


Figure 31: Function To conduct CEI (in its two versions) represented in the OF framework. In its atrial manifestation (left-hand) this function is *realized by* means of the process of Depolarization of CSA myocytes. In its ventricular manifestation (right-hand), it is in turn *realized by* means of the Depolarization of CSV myocytes.

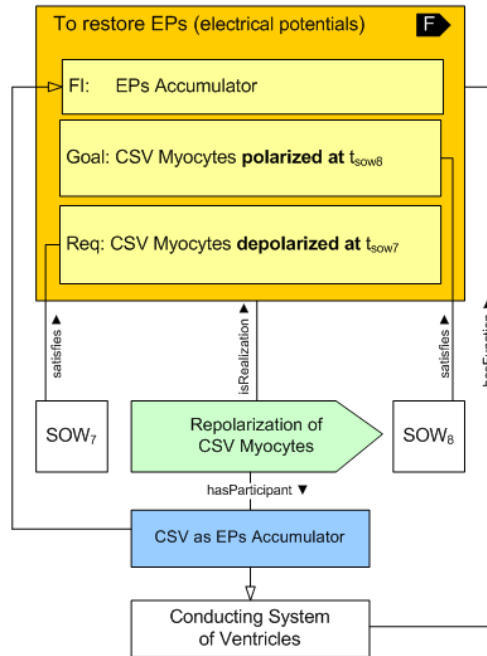


Figure 32: Function To restore EPs represented in the OF framework. This function is *realized by* means of the process of Repolarization of CSV myocytes. Speaking of the requirements for state of the world SOW<sub>1</sub> to be reached, we could say that as a matter of fact before their repolarization the ventricles have actually done with the contractions they make. However, we have not considered such thing as a requirement for not committing to associations between the heart mechanical and electrical activities which go beyond our scope.

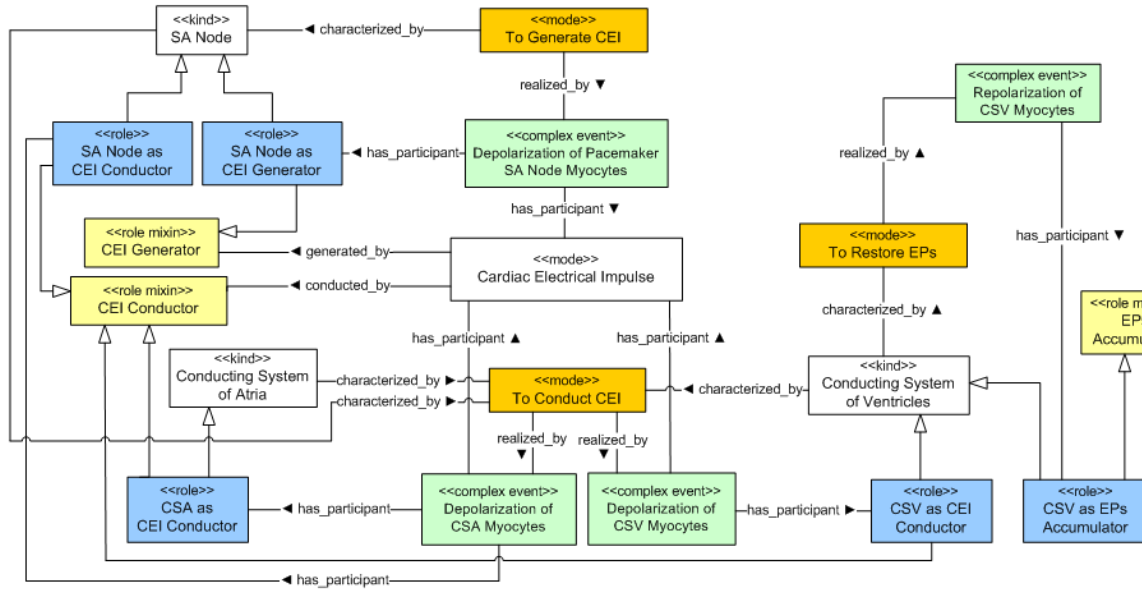


Figure 33: Model of heart electrophysiological functions. The colors have no semantics in the OntoUML abstract syntax, but are used here to provide a visual association to the OF schemata presented previously. The SA node as CEI conductor *participates in* the latter process, since the SA node is a *part of* the CSA which is actually fundamental in the whole conduction thing. Albeit not illustrated, the process of Depolarization of Pacemaker myocytes is *part of* Depolarization of CSA myocytes. It is also worthwhile to highlight that the relation *participates in*, inverse of *has participant*, holds for every assertion of the latter.

CSA myocytes or Depolarization of CSV myocytes. Finally, the function To restore EPs (on the right) also *characterizes* the conducting system of ventricles. This function is *realized by* the process of Repolarization of CSV myocytes. The four processes appearing here are *complex events* since at least two sub-events can be distinguished in each of them, viz. those two which are divided by the time the electrical current reaches its higher value (in a graphic, its peak) at. Except by *generated by* and *conducted by*, all the relations asserted in this model have their inverse counterparts holding as well. Finally, the following relations hold between the functions aforementioned: To generate CEI *enables* the atrial manifestation of To conduct CEI. The latter then *enables* its ventricular manifestation, which in turn *enables* To restore EPs.

The class-level *has participant* relation as used in the model depicted in Figure 33 is formally described without time argument as follows - it is built on top of its instance-level version defined in Table 1.

$$\text{has\_participant}(P, C) =_{\text{def}} \forall p ( \text{instance\_of}(p, P) \rightarrow \exists t, c ( \text{occurring\_at}(p, t) \wedge \text{instance\_of}(c, C, t) \wedge \text{has\_participant}(p, c, t) ) )$$

Intuitively, a process has a given continuant as participant (playing some role) iff it participates in some time instant the process is occurring at. This is in consonance with the OBO Relation Ontology assignment. The inverse *participates in* holds as well for all *has participant* assertions in the model of Figure 33.

As it can be noticed, we make use of OF's *realized by* relation (and its inverse *is realization*) as well as the *enables* relation holding between functions. For a formal account of these, we report to OF (114), since their definitions take many other notions in consideration which fall outside our scope. Besides, we introduce here some germane relations not present in OBO RO (e.g., *generated\_by*, *conducted\_by*), but that are built on top of RO's basic relations. As follows, we define those relations in order to restrict their interpretation. However, before we can define what it means to state that a continuant has been generated by another, we need to define the notion

of production. The instance-level relation *produced\_by* holds between a continuant and a process. As formally described below, a continuant  $c$  is produced by a process  $p$  iff there exists one and only one time instant  $t_1$  such that  $t_1$  is the last instant of  $p$ ,  $p$  has  $c$  as participant at  $t_1$ , and for all time instants  $t$  earlier than  $t_1$  then  $c$  does not exist at  $t$ . The class-level relation *produced\_by* is defined subsequently.

$$\mathbf{produced\_by}(c, p) =_{def} \exists t_1 ( \mathbf{last\_instant}(p, t_1) \wedge \mathbf{has\_participant}(p, c, t_1) \wedge \forall t ( \mathbf{earlier}(t, t_1) \rightarrow \neg \mathbf{exists}(c, t) ) )$$

$$\mathbf{produced\_by}(C, P) =_{def} \forall c ( \mathbf{instance\_of}(c, C) \rightarrow \exists p ( \mathbf{instance\_of}(p, P) \wedge \mathbf{produced\_by}(c, p) ) )$$

Notice that to state that a continuant participates in some process entails it exists during that process, cf. Table 1. We are now able to proceed by giving a definition for the notion of generation. A continuant  $c$  is generated by another continuant  $c_1$  iff there exists some process  $p$  such that, for all time instants  $t$  at which  $p$  is occurring then  $p$  has  $c_1$  participating as an agent, and  $c$  is produced by  $p$ . See also the class-level version.

$$\mathbf{generated\_by}(c, c_1) =_{def} \exists p ( \forall t ( \mathbf{occurring\_at}(p, t) \rightarrow \mathbf{has\_agent}(p, c_1, t) ) \wedge \mathbf{produced\_by}(c, p) )$$

$$\mathbf{generated\_by}(C, C_1) =_{def} \forall c ( \mathbf{instance\_of}(c, C) \rightarrow \exists c_1 ( \mathbf{instance\_of}(c_1, C_1) \wedge \mathbf{generated\_by}(c, c_1) ) )$$

The notion of conduction, in turn, is a bit more complex. First, following UFO we take the meta-category *mode* into account. The reason is that an entity which is object of conduction, like the cardiac electrical impulse (CEI), needs to *inhere in* some conductor in order to exist (34, Chapter 6). Thus, it is said to be *existentially dependent* on some conductor. The CEI is modeled here as a mode, just as a symptom is, which only exists by inhering in some patient. Before delimitating what does conduction mean, we present below the instance-level primitive relation of *inherence*, together with its correlated class-level relation of *characterization*. Inherence is an irreflexive, asymmetric and intransitive type of existential dependence relation; characterization can only be applied if  $F$  (see the formulae below) is an instance of the UFO meta-category *moment universal* (from which *mode* is a specialization). In this case, we add the restriction that the variable  $F$  ranges over functions (a specific type of mode).

$$\mathbf{Irreflexivity:} \quad \forall c, t \quad \neg \mathbf{inheres}(c, c, t)$$

$$\mathbf{Asymmetry:} \quad \forall c, c_1, t \quad \mathbf{inheres}(c, c_1, t) \rightarrow \neg \mathbf{inheres}(c_1, c, t)$$

$$\mathbf{Intransitivity:} \quad \forall c_1, c_2, c_3, t \quad \mathbf{inheres}(c_1, c_2, t) \wedge \mathbf{inheres}(c_2, c_3, t) \rightarrow \neg \mathbf{inheres}(c_1, c_3, t)$$

$$\mathbf{Exist. dependence:} \quad \forall c_1, c_2 \exists t_1 \mathbf{inheres}(c_1, c_2, t_1) \rightarrow \forall t ( \mathbf{exists}(c_1, t) \rightarrow \mathbf{exists}(c_2, t) \wedge \mathbf{inheres}(c_1, c_2, t) )$$

$$\mathbf{characterized\_by}(C, F) =_{def} \forall c, t ( \mathbf{instance\_of}(c, C, t) \rightarrow \exists f ( \mathbf{instance\_of}(f, F, t) \wedge \mathbf{inheres}(f, c, t) ) )$$

The inverse class-level *characterizes* holds for all function universals and their realizer substantials represented in the model of Figure 33.

We can then proceed to formally describe the relation conducted by between two continuants  $c$  and  $c_r$ . This

relation is characterized here using the three formulae below. The first of these formulae states that if  $c$  is conducted by  $c_r$  then there is a (conduction) process  $p$  that eventually occurs and that, in all instants that this process occurs, both  $c$  and  $c_r$  participate in this process. Moreover, the formula states that  $c$  inheres in  $c_r$  during this entire process and only during this process. Putting this formula together with the condition of existential dependence for the inherence relation defined above we have that participating in this conduction process is an essential condition for  $c$ .

$$\begin{aligned} \text{conducted\_by}(c, c_r) \rightarrow & \exists p, t_1 \text{ occurring\_at}(p, t_1) \wedge \forall t ( \text{occurring\_at}(p, t) \rightarrow \text{has\_participant}(p, c, t) \\ & \wedge \text{has\_participant}(p, c_r, t) ) \wedge ( \forall t_2 \text{ inheres}(c, c_r, t_2) \leftrightarrow \text{occurring\_at}(p, t_2) ) \end{aligned}$$

The next formula states that in all instants that  $c$  inheres in  $c_r$  (i.e., all instants that  $c$  exist),  $c$  occupies a spatial region  $r_1$  that is a proper part of the spatial region  $r$  occupied by its bearer (the conductor). Moreover, the formula states that given a time instant  $t$ , there is only one region occupied by  $c$  in that instant (analogously for the conductor  $c_r$ ). Finally, the formula (indirectly) states that during the conduction process  $p$  (i.e., during the lifetime of  $c$ ),  $c$  occupies all proper parts of  $r$  but also that no proper part of  $r$  is occupied by  $c$  more than once during the process  $p$ .

$$\begin{aligned} \text{conducted\_by}(c, c_r) \rightarrow & \\ \forall t ( \text{inheres}(c, c_r, t) \rightarrow & \exists r, r_1 ( \text{located\_in}(c_r, r, t) \wedge \text{located\_in}(c, r_1, t) \wedge \text{part\_of}(r_1, r) \wedge \\ \forall r_2, r_3 ( \text{located\_in}(c_r, r_2, t) \wedge & \text{located\_in}(c, r_3, t) \rightarrow (r_2 = r) \wedge (r_3 = r_1) ) \wedge \\ \forall r_4 ( \text{part\_of}(r_4, r) \rightarrow & \exists ! t_1 \text{ inheres}(c, c_r, t_1) \wedge \text{located\_in}(c, r_4, t_1) ) ) ) \end{aligned}$$

Finally, the following formula states that given any two instants  $t_1$  and  $t_2$  such that  $c$  inheres in  $c_r$  both in  $t_1$  and  $t_2$  and that  $t_1$  is the instant immediately earlier  $t_2$  then in each of these instants,  $c$  occupies regions adjacent to each other.

$$\begin{aligned} \text{conducted\_by}(c, c_r) \rightarrow & \forall t_1, t_2, r_1, r_2 ( \text{inheres}(c, c_r, t_1) \wedge \text{inheres}(c, c_r, t_2) \wedge \text{located\_in}(c, r_1, t_1) \\ & \wedge \text{located\_in}(c, r_2, t_2) \wedge \text{immediately\_earlier}(t_1, t_2) \rightarrow \text{adjacent\_to}(r_1, r_2) ) \end{aligned}$$

Confer below the relation of *immediately\_earlier* holding between two time instants.

$$\text{immediately\_earlier}(t_1, t_2) =_{\text{def}} \text{earlier}(t_1, t_2) \wedge \neg \exists t ( \text{earlier}(t, t_2) \wedge \text{earlier}(t_1, t) )$$

A class-level version of the *conducted\_by* relation can then be defined as follows.

$$\text{conducted\_by}(C, C_r) =_{\text{def}} \forall c \exists t \text{ instance\_of}(c, C, t) \rightarrow \exists c_r \text{ instance\_of}(c_r, C_r, t) \wedge \text{conducted\_by}(c, c_r)$$

At this point our ECG theory has already gained some substance. We can then finally focus on the ECG itself, but now with an established background.

## 5.4 The Electrocardiogram

The contents of this section are based on (129, 22, 128). The models presented in what follows are based on evidence present in the medical textbooks just mentioned but also synthesize concerns present in current ECG standards (leaving out technological aspects). Once we have set the ground of anatomy and physiology, we can concentrate our ontological analysis in the ECG properly. The ECG (in German, the elektrokardiogram, EKG) was probably the first diagnostic signal to be studied with the purpose of automatic interpretation by computer programs (22). The reason for such an interest in computing ECG records is that the analysis of the ECG waveform can help

to identify a wide range of heart illnesses, which are distinguished by specific modifications on the ECG elementary forms. The ECG is indeed the most frequently applied test for measuring the heart activity in Cardiology. In comparison with other examination procedures, it is fast, cheap and non-invasive.

Let us start our study on the side of the patient. The ECG Record<sup>12</sup> is acquired from a given Patient by a Recording device in the context of a Recording session, see Figure 34. The record is then *produced by* such a session in the precise sense formally described in the previous section. As the session unfolds in time from a given start to an end date/time, the latter indeed determines the end of the production of the record. Such an ECG record can be part of the patient's EHR, but as this may not be the case we prefer not to assert a parthood relation for that. The ECG record has as an *essential part* an ECG Waveform, which is elaborated further on.

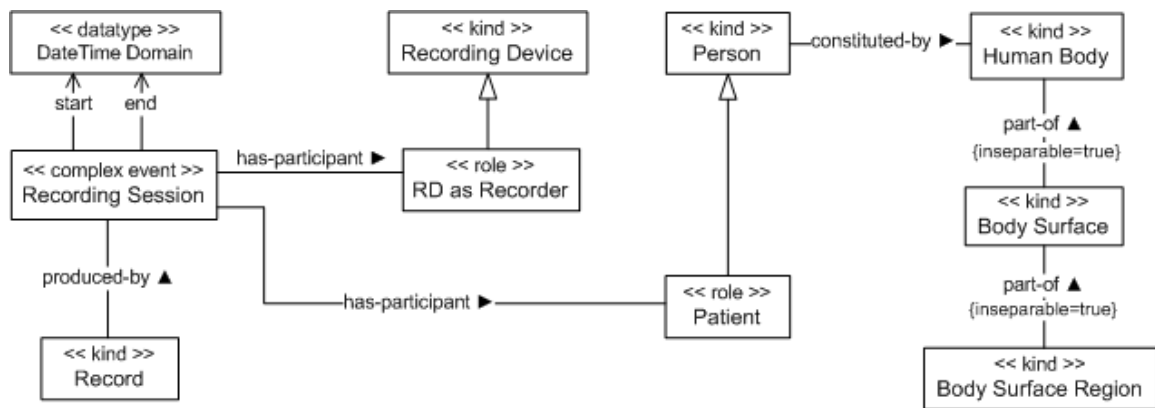


Figure 34: Model of the ECG recording session context. The ECG Record is *produced by* a Recording session that has a start and end date/time. They are two properties of the session which are projected into the Date Time *quality domain* - cf. (34, Chapter 6). A session has as participants an RD as recorder (role, type of Recording device) and a Patient. The latter is a *role* played by a given Person, who is *constituted by* a Human body. Although very basic, this is in fact worth to mention since Body surface region is the entity which is object of measure for the ECG acquisition; and that comes to bridge the ECG sub-ontology to that of anatomy.

Notice that the ECG recording session is an example of complex event. Indeed, (many) observations (or measurements, loosely speaking) are made by the recording device at and in between the session's temporal boundaries, see Figure 35. These observations (atomic events) are actually what allows the ECG to gradually take form. They are evenly spaced in time, forming then an *Observation series* that respect a non-zero *period* of time. The *sample rate* of the ECG Waveform accounts for the inverse of that period. Sample rate values are projections into the conceptual space of Hertz (Hz, samples/second) by considering the period in seconds.

The observations are meant for measuring electrical potential differences (p.d.) around the patient's body surface with the result of producing samples. Every observation produces an electrical potential sample (in the scale of millivolts). The sample values are projections into the voltage conceptual space, i.e., that of the real numbers. We submit that every sample is a *grain of* a *Sample sequence* - an ordered collective of such samples. The decision of assigning a sample sequence to the collective type is based on the following rationale: (i) the usual number of samples of such a sample sequence in a typical ECG record is to an extent of thousands; (ii) few samples can even be lost with no damage to the whole, e.g., if the device for some reason jumps an observation, the last sample value can be repeated with no significant impact (the waveform is usually dense, with sample rates of 256, or 300 Hz). A sample sequence is a projection to the conceptual space of an ordered sequence (in the mathematical

<sup>12</sup>Note that although the term "record" is very general even within Health Informatics in particular, we refrain from assigning a rubric like "ECG record" to avoid verbosity. This is because, if that were the case, many classes in our domain would require for the "ECG" prefix as well. Nevertheless, each class is accompanied by a unique ID with the "ecg" signature before it, cf. Section 5.7.2, and this will suffice for us.

sense) of real numbers (standing for p.d. values).

We draw attention here to the relation *grain of*, which uses a different term for the same thing referred to in UFO as “member of” (34, Section 5.5). We have abided by the first designation in order to comply with the terminological usage in the biomedical literature. In (95), the notion of *grain of* is treated by Rector et al. with the same intent we have here. In consonance with the UFO ontology, *grain of* is a specific type of *parthood*, but a non-transitive one. Besides intransitivity, it stands for irreflexivity, asymmetry, and weak supplementation introduced above. Although transitivity does not hold between two *grain of* relations, transitive holds if *grain of* is followed by *subcollection of* (34). For example, if cellulose is *grain of* trees, and trees are *grain of* forests, then it follows that cellulose is not *grain of* forests. On the other hand, since a SA node pacemaker myocyte is *grain of* (a collective of) SA node pacemaker myocytes, and SA node pacemaker myocytes is *subcollective of* SA node myocytes, then it follows that a SA node pacemaker myocyte is *grain of* SA node myocytes. The notion of *sample sequence of* that relates a sample sequence to an observation series is formally described in the next section.

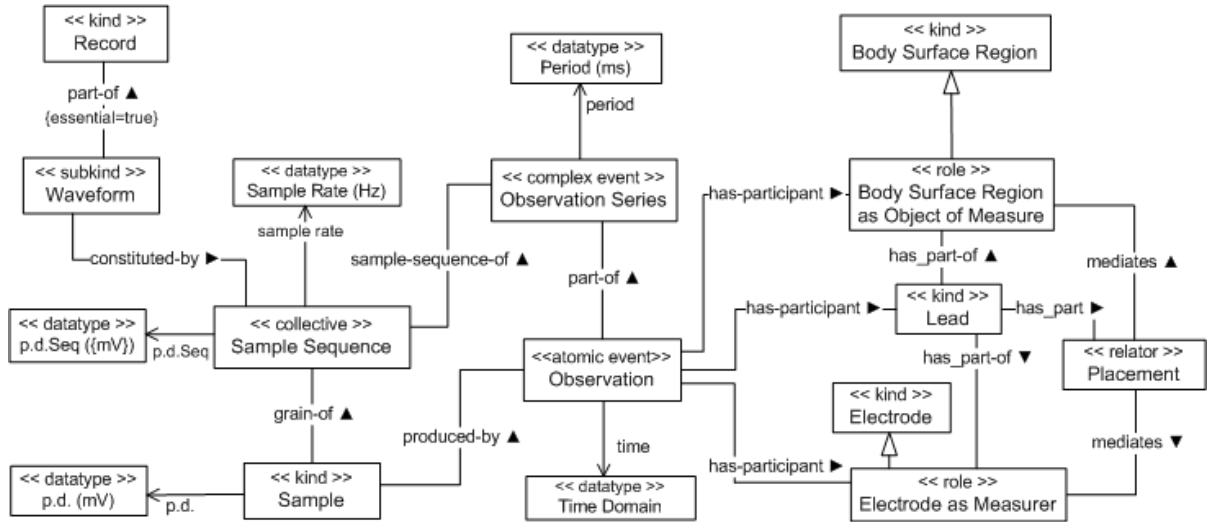


Figure 35: Model of the ECG acquisition mechanism. Though not illustrated here, an *Electrode* is *part of* a *Recording device*. The kind universal from which the *subkind* *Waveform* takes its identity principle is not shown in this model neither, but in the model that follows. The *participates in* relation which is inverse of *has participation* holds towards *Observation* for the *Electrode* as measurer and the *Body surface* as object of measure, but not for *Lead*. We draw special attention here to the *relator* *Placement* (on the right) mediating the *roles* *Electrode* as measurer and *Body surface* region as object of measure. The relator is somewhat that, according to UFO, connects the *qua individuals* correspondent to these role entities. Thus, as postulated in UFO, these *exist in* the placement.

Let us now focus on the way an observation is made by means of a pair of electrodes (each *Electrode* is *part of* the *Recording device*), see the right-hand of the model in Figure 35. As discussed in the previous section, those electrical potentials manifested on the body surface are result of the heart electrical activity, namely, of the cardiac electrical impulse that even though more intense on the heart, reaches almost all regions of the body surface (e.g., the left ankle, the right and left wrists)<sup>13</sup>. From the time this was discovered on, many advances over the years led to the practice of Electrocardiography. It is the technique of recording the electrical signal generated by the heart activity. The mechanism is basically the following. By means of two electrode placements

<sup>13</sup>Notice, as might be expected, that any potential differences within the body can have an effect on the electrical potentials detected on the ECG. Movements require the use of skeletal muscles, which then contribute to the changes in voltages detected using electrodes on the body surface. For this reason, the ECG is distinguished with respect to the state of the patient when it is being acquired. In the “resting ECG” - the standard one, which is considered here, the patient should be essentially motionless, i.e., he/she should remain as still as possible for not influencing on the diagnostic (129).

(cf. **Placement** in Figure 35) on two specific regions of the patient's body surface, the recording device performs such an observation. It is supposed to measure the p.d. between these two electrode placements. However, since measuring the p.d. between two points provides only a partial point of view of the heart activity, usually multiple observations are made at the same time instant to capture a multiple view of the heart activity. While a series of observations of the p.d. between the same two electrode placements denotes an **Observation series**, multiple series that share the same structure in the time axis (i.e., the same beginning, end and period) denote a *correlated observation series*. But, for us, the main point is that: each of those viewpoints that emerge from one single **Observation series** defines an **ECG Lead**. In summary, a lead is a viewpoint of the heart activity that emerges from an observation series of the p.d. between two electrode placements on specific regions of the patient's body surface, see Figure 36. Put together, multiple leads provide an accurate picture of the heart behavior. From a metaphysical standpoint, we consider as being meta-properties of **Lead** those of *rigidity*, *external independence* and the *provision of an identity principle*. It is therefore assigned to the *kind* type.

One might notice by looking at the model in Figure 35 the presence of a new (in this text) universal type, viz., the *relator* (34, Chapter 6). It is contemplated in the UFO ontology for capturing those real-world entities that connect at least two continuants, as a kiss, a handshake, the enrollment of a student in some educational institution, a covalent bond, and especially, for us, a **Placement**. The latter is the physical contact between an *electrode* and a specific *body surface region* to measure a voltage value. A quotation from Guizzardi (34, p. 240) tells us that, a relator individual is “an aggregate of all *qua* individuals that share the same foundation”. So, if Bill kisses Monica, the individual “Bill and Monica's kiss” is indeed the mereological sum of “Bill *qua* kisser” and “Monica *qua* kisser”. By the same token, the individual “placement of electrode *x* on the surface of the right wrist of patient John Doe” denotes the mereological sum of “electrode *x qua* measurer” and “surface of the right wrist of patient John Doe *qua* object of measure”. In general, let *x*, *y* and *z* be three distinct individuals (from the meta-level standpoint) such that: (a) *x* is a relator individual; (b) *y* is a *qua* individual and *y* is part of *x*; (c) *y* inheres in *z*. In this case, we say that *x mediates z* (34). Formally, we have that

$$\forall x, y \text{ mediates}(x, y) \rightarrow \mathbf{Relator}(x) \wedge \mathbf{Continuant}(y)$$

$$\forall x \mathbf{Relator}(x) \rightarrow \forall y ( \text{mediates}(x, y) \leftrightarrow ( \exists z \mathbf{QuaIndividual}(z) \wedge \mathbf{part\_of}(z, x) \wedge \mathbf{inheres}(z, y) ) )$$

$$\forall x \mathbf{Relator}(x) \rightarrow \exists y, z ( (y \neq z) \wedge \text{mediates}(x, y) \wedge \text{mediates}(x, z) )$$

Recall that predicates “relator”, “continuant” and “qua individual”, if considered in a domain-specific representation, would work actually as higher-order predicates, just as all the others we have been using thus far (viz., category, kind, phase, mode and so on). The relation *mediates*, as impinged by the formulae above, always holds between a relator and a continuant; moreover, it stands for *irreflexivity*, *asymmetry*, *intransitivity* and *existential dependence* on at least two individuals.

By shifting to the physician's perspective, we shall put in focus the objects of ECG analysis. Heart beats are mirrored to cardiac cycles that compose the **ECG Waveform**, see Figure 37. There are two main characteristics for interpretation in the ECG: (i) the morphology of waves and complexes which compose a cardiac **Cycle**; and (ii) the timing of events and variations in patterns over many beats. In this way, the analysis of the ECG waveform supports identifying a wide range of heart diseases. The characterization of each cardiopathy manifests itself by specific modifications on characteristics (i) and (ii).

A canonical **Cycle**, as introduced by Dutch physiologist Willem Einthoven, has waves named PQRST. They are outlined as **P wave**, the mereological sum of the Q, R and S waves (so-called **QRS complex**), and **T wave**.

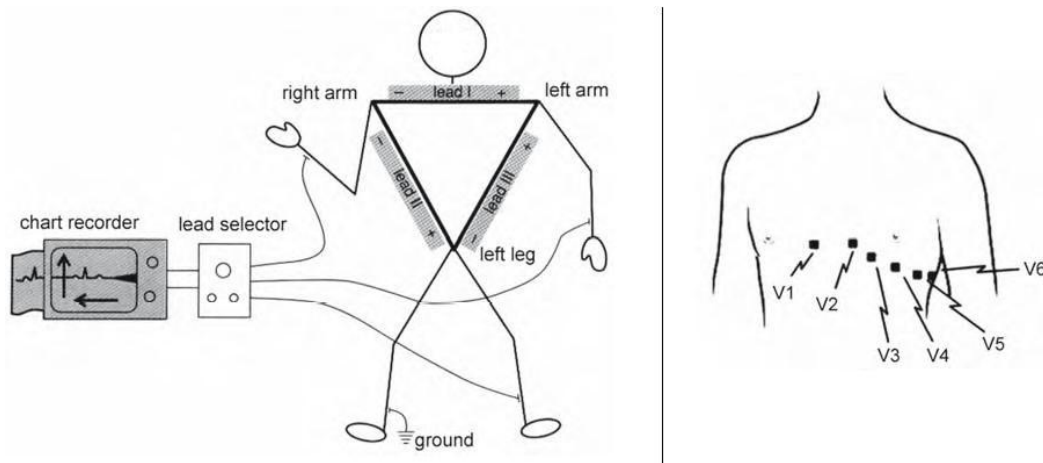


Figure 36: ECG leads - adapted from (129). Electrocardiography has standardized twelve leads which provide viewpoints for analyzing the heart activity. Some of these are bipolar leads as they make use of a single positive and a single negative electrode between which electrical potentials are measured. They are the so-called *limb leads* (see the picture on the left), and are obtained by measuring the p.d. between (i) the surfaces of the left and right arm (usually on the wrists), (ii) the surfaces of the left leg and right arm, and (iii) left leg and left arm respectively. Arm and leg placements are usually made on the wrist and the ankle, respectively. The other nine leads are unipolar, as they use only a single positive electrode and a configuration of the other electrodes to serve as a composite negative electrode. Three of them are the *augmented leads*, viz., aVL, aVR and aVF. Each of them is obtained by opening the correspondent resistor (in the wire) constituting one of the three limb leads. Finally, six *precordial*, or *chest leads* (see the picture on the right), viz., V1, V2, V3, V4, V5 and V6, comprise six electrode placements at the rib cage near the heart. They also make use of weighting resistors, but in combination with a common reference electrode placement at the chest (not shown in the picture) called the Wilson central terminal (WCT). Every chest lead is then obtained by the p.d. between one of those electrode placements and the WCT. When an ECG is recorded some (or even all) of those twelve leads are recorded simultaneously (or not) at different device channels. Due to time constraints, in the ECG ontology we are developing here we have not represented the effect of multiple leads put together.

To anticipate the next section, the P wave and QRS complex map the depolarization of atria and ventricles, respectively. The atrial and ventricular myocardial contractions start normally at the offset of these waves. At last, the T wave maps the repolarization of ventricles.

The basic entity, the substantial sortal which is the center our analysis here is ECG form. We are referring, precisely, to the form that emerges from a given sample ordered collective, i.e., which is constituted by it. Although this is mostly presented (especially in the ECG domain) in a graphical, visual way, the form itself is a more general entity. It is the emergent pattern denoted by the connection of the adjacent (2-tuple) time-voltage values of a given sequence<sup>14</sup> in their conceptual, or geometric space. So, the forms we are talking about are like the pattern arisen by the connection between 3-tuple vertex values of the Great Pyramid of Giza when projected into some three-dimensional conceptual space; or between 2-tuple values of latitude-longitude that define the (fiat) territorial area of the Amazon in the geographic coordinate system. Ergo, the ECG form under consideration here is a type of the *category* Geometric form. However, the specific notion of geometric form employed in the ECG domain is *time-series*, since it presupposes a bi-dimensional space whose horizontal axis is denoted by the time dimension. The vertical axis in turn stands for the projection of p.d. (voltage) values into the conceptual space of the real numbers. Note that a consequence of this is that any geometric operation (say, reflection in relation to the vertical

<sup>14</sup>Notice that the observations are carried out periodically over the time, giving then rise to a sample sequence with a particular sample rate. So, from both values of the start time of the ECG recording session and the sample rate, we easily have the time value for each sample value composing then such 2-tuples. We therefore have not represented the sequence an ECG form is constituted by as a 2-tuple sequence, but only as a sample sequence for the sake of simplicity.



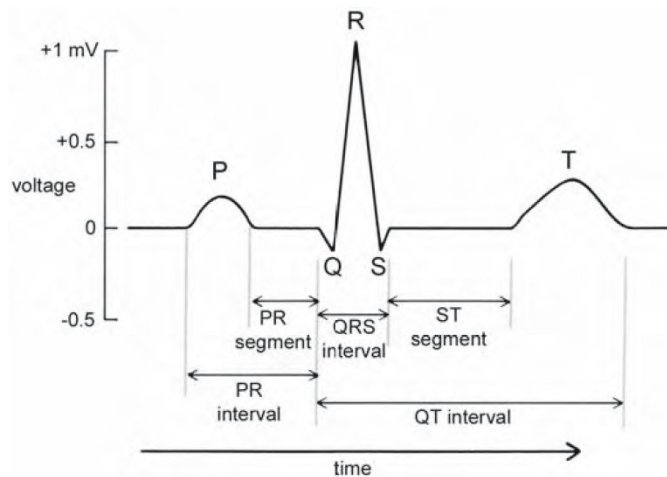


Figure 37: A typical ECG waveform for one cardiac cycle measured from the lead II (the most referred one). The P wave denotes atrial depolarization, the QRS complex indicates ventricular depolarization, and the T wave denotes ventricular repolarization. The end of the T wave connects to the beginning of the next P wave to mark the beginning of the next cycle. The notion of interval is not considered in our theory since, not being part of the ECG waveform, it falls outside our scope (cf. Subsection 5.7.1). Any *fiat* interval demarcation can, however, be carried out in the ECG waveform, including the referred PR and QT intervals. Source: (129, p. 192).

axis) on a given instance of form transforms it into another one. At last, every form in an ECG waveform bears the property of having a *duration* - normally attributed in milliseconds.

A given ECG Form can be either elementary or not. Every ECG Elementary form is, as the name suggests, elementary for ECG analysis. These forms are those patterns appearing (in a canonical reading) in every cardiac cycle (including the cycle itself) that either directly map a cohesive and germane electrophysiological event in the heart behavior or connect two forms which do it<sup>15</sup>. Examples of it are the waves and segments depicted in Figure 37. An Non-Elementary form in turn is any arbitrary ECG form which is not an elementary one. An example of such ECG form is that, loosely speaking, constituted by just a part of the T wave, or by the right half of the P wave and the left half of the PR segment. The ECG Waveform is the ECG form *constituted by* the whole Sample sequence resulting from the full Observation series carried out at and in between the session's time boundaries. In other words, the ECG form resulting from an ECG Recording session. The waveform itself is also a non-elementary form, as it as a whole is not (an elementary) object of physician's analysis. Elementary forms in turn are of different natures. Namely, Wave, complex (only the QRS complex), (line) Segment and Cycle. This partition is complete, and is made under the differentiae of specific geometric properties that are elaborated further on in this text. Another partition finds place for elementary forms, viz., the disjoint and complete distinction between Normal and Abnormal elementary forms. This partition is carried out for each type of elementary form denoted by the former partition, but under the differentiae of whether or not a given elementary form individual matches the canonical geometric pattern for the type it instantiates.

Before we proceed in our analysis to take the subtleties that define a cardiac cycle and the waves, complexes and segments it is composed of into account, let us justify why they are all *subkinds* of Elementary form which is in turn assigned under the rubric of the ECG Form *kind*. For supporting our analysis, consider the piece of ECG waveform depicted in Figure 37. The first question that may be raised in looking at that ECG waveform is: if that is an instance of ECG waveform, what are the (other) instances we have at hand? For example, if one says

<sup>15</sup>For instance, the P wave maps the depolarization of atria, while the QRS complex maps the depolarization of ventricles; but the PR segment connects the former to the latter. We address this matter in detail in the next section.

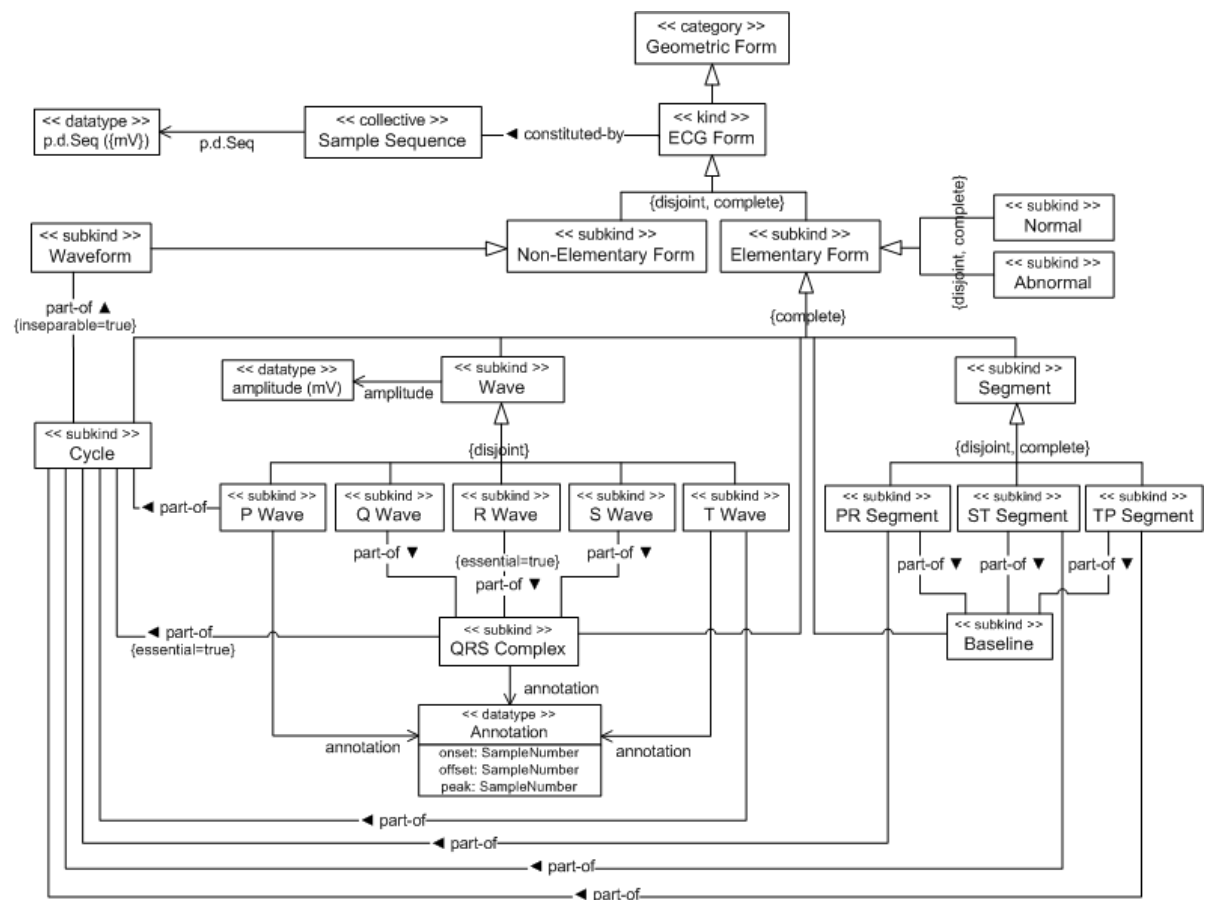


Figure 38: Model of the ECG waveform (on the side of the physician). This entity stands for a *constitutes* relation (inverse of *constituted by*) with ECG form as range. Notice that what connects it to the model on the side of the patient is the entity Sample Sequence. Indeed, with no record of the ultimate data the exam brings in, the physician is not able to analyze anything. On the other hand, only by possessing such ultimate data (no matter the particular way it is presented), the physician has the necessary material for playing his/her clinical role. By ultimate data we mean the sample values of the electrical potentials measured on regions of the patient's body surface over time. The physician analyzes the ECG by considering cycles. Each one represents a heart beat. The classes Normal and Abnormal under the genus Elementary Form are elaborated in the next section. A disjoint assertion holds between only a Wave and a Segment. A *has part* relation holds between the Cycle and the QRS complex, and between the latter and the R wave.

“the duration of that QRS complex is 40 ms”, or “that P wave looks normal”, to what things “that QRS complex” and “that P wave” refer? To put it differently, what principle allows us to distinguish “that QRS complex” from “that P wave”? Indeed, these instances are, above all, *ECG forms* - in the sense elaborated above. The reason is that the notion of ECG form provides its instances with an identity principle, viz., two ECG forms are the same if their sequences of 2-tuple values in the time vs. voltage conceptual space coincide. In contrast, the classes Elementary form, Wave, QRS complex, Segment or Cycle fall short in providing such a principle, but rather, their instances carry a unique identity as instances of ECG forms. They are in fact subkinds of ECG forms derived by two differentiae: (i) for the partition between elementary and non-elementary forms, the criterion used is whether or not the ECG form either directly maps a cohesive electrophysiological event in the heart behavior or connects two forms which do it; (ii) for the partition between wave, complex, segment and cycle (all of them are elementary forms) the criteria used are related to specific geometric properties (e.g., does it have a peak?) their instances bear as instances of ECG forms. Notice that classifying a given entity into any of these universals is only possible by taking account of the notion of (the substantial sortal) *ECG form*. We reflect on this whole example to,

again, draw attention to the role of ontological principles as providing support for our decision. In fact, at a first glance, a clear understanding of what those entities are was far from being trivial. The best practice of “looking for the substantials first of all for defining the backbone taxonomy”, however, as suggested in OntoClean, has been worthwhile.

Accordingly, we submit definitions for wave, complex, segment and cycle as follows. Notice, however, that we do not get that deep in their *interpretation*, as this is the purpose of the next section. A **Wave** *is a* Elementary form that bears necessarily the property of having a peak. For instance, one is able to recognize the peaks of the P, Q, R, S and T waves in Figure 37 as being defined by their point of higher y-coordinate value in module. Every wave has a voltage *amplitude* (usually in the scale of millivolts), i.e., the projection of its peak sample value into the p.d. conceptual space. The subtypes of waves just mentioned are all those we consider here<sup>16</sup>.

A complex (from which the QRS complex is the only representant) in turn *is an* Elementary form that can be composed of more than one wave. The QRS complex (in the canonical reading) is the mereological sum of the Q, R and S waves. The QRS complex as well as some of the waves (viz., P and T waves) are usually annotated in the ECG waveform. That is, their onset (beginning), peak and offset (end) time values are marked, say, by a computer program such that it can be better visualized by a physician when carrying out an ECG analysis. The peak of the QRS complex coincides to the peak of the R wave. Again, as the ECG is a periodic waveform, these points are often annotated just by storing their sample numbers in the sample sequence, cf. the model in Figure 38.

Finally, a (line) **Segment** *is an* Elementary form that connects two waves and does not have a peak. Subtypes of segments are the PR segment<sup>17</sup>, ST segment and TP segment. This term “TP segment” is not actually referred to in the ECG jargon. The entity we bring in by it is the segment that connects the offset of each T wave to the onset of the next (if there is) P wave. Such an entity is often found in the literature under the rubric of “Baseline”. We, however, refrain from committing to the latter term for denoting that entity. The reason is that **Baseline** is in essence something else, viz., all the ECG form (even discontinued, or interrupted by waves) with a null voltage value that is eventually composed by every segments in the whole ECG waveform. “Baseline” (even metamorphosed as “Isoelectric line”) indeed exemplifies ambiguity in the ECG domain, since it is a term used to mean both entities aforementioned.

Overall, the **Baseline** (as denoted here) indicates the absence of electrical activity in the heart. This roughly coincides to the mereological sum of the recording at and in between the time windows where the heart is in its resting state. The genuine ECG elementary forms are referred to as giving a picture for some particular activity of the heart. The baseline, contrariwise, does not indicate any activity of the heart, but actually the absence of heart activity. Nonetheless, is not this also an important information? Indeed, it is the only information that can be known from an ECG, let us say, when a patient is not alive anymore. Ergo, we submit that it does is a germane entity in the context of the ECG waveform. More than that, we argue that every ECG waveform part has a meaning only in the context it is in between, as we are dealing with a periodic waveform pattern. Hence, the act of cutting off any part of the waveform (e.g. the baseline) would alter by far the whole thing.

An additional aspect of segments to be mentioned is that even though not directly, they can be said to possess onset and offset points as well. These can be derived always from the offset of the wave that precedes the segment and from the onset of the wave that follows it.

<sup>16</sup>There is also a U wave, which nature is however uncertain in the ECG domain literature. For that reason, the property *complete* does not appear in the partition of types of waves in the model of Figure 38.

<sup>17</sup>Also referred to as “PQ segment”, as in Figure 37. This is because, depending on the ECG lead in hand, the Q wave is not detectable.

We then have covered all the parts that compose a **Cycle**. Overall, the combination of the P wave, PR segment, QRS complex, ST segment, T wave and TP segment composes the cardiac cycle. It should be clear, however, that this is a representation of a canonical cardiac cycle, which is usually said to be better approached by lead II - viz., the most referred one for canonical study. In practice, there are cycles with missing ECG elementary forms (e.g., a missing Q wave) for several reasons, e.g., because it is not well visible from the ECG lead in hand. The cardiac cycle is of significant importance as a unit also for the calculation of the heart rate. This can be easily obtained by taking the inverse of the average period of two or more cardiac cycles. At last, normally many cardiac cycles compose an ECG **Waveform**. Actually, an ECG *form* even shorter than a cycle could be said an ECG waveform. For instance, consider the starting of an ECG session that is promptly canceled for some reason. However, such a situation is very odd and would most likely lead to the annulation of such exam.

## 5.5 Basic ECG Interpretation

Once all that is ontologically understood, we are able to address the structure of the ECG cardiac **Cycle** in terms of the meaning the elementary forms bring in. The cardiac activity begins with the firing (excitation) of the SA node in the right atrial myocardium. This firing, however, is not detected by standard ECG because the number of SA myocytes is not enough to create electrical potentials detectable on the body surface, i.e., with a high enough amplitude to be recorded with distal electrodes (signal amplitude is lost as it dissipates through the conductive medium). Thus, the first deflection that takes place in the ECG cycle is actually a result of the atrial depolarization; it is the so-called **P wave**. This represents the coordinated depolarization of the right and left atria and indirectly indicates the onset of atrial contraction. The P wave is normally around 80 - 100 ms in duration. As the P wave ends, the atria are completely depolarized and are beginning contraction.

The ECG signal then returns to the y axis origin and stays there from the P wave offset to the Q wave onset (to the R wave onset, if the Q wave is not present). That characterizes the PR segment, which corresponds to the spreading of the CEI (not large enough in amplitude to be detected in the ECG) to the AV node and AV bundle. These structures then slow the CEI as the ventricles are filled in with blood. Roughly, 160 ms after the beginning of the P wave, the right and left ventricles begin to depolarize, resulting in the QRS complex. Typically, the first negative deflection is the Q wave, the large positive deflection is the R wave, and if there is a negative deflection after R wave, it is the S wave. The exact shape of the QRS complex actually depends on the ECG lead in hand. The QRS complex offset indirectly indicates the beginning of ventricular contraction, which varies between 60-100 ms (usually 80 ms) in duration.

At the time of the QRS complex onset, atrial contraction has normally ended, and the atria are repolarizing. However, the effect of the atria repolarization is sufficiently masked by the much larger amount of tissue involved in ventricular depolarization occurring at the same time and is thus not detected in the ECG waveform. Then, from the QRS complex offset to the T wave onset, the ECG signal stays neutral in amplitude while ventricles are doing contraction. This is the ST segment, which though does not map (directly) anything, is a very important input for diagnostic of the myocardial ischemia. The ventricles go through repolarization after contraction, what gives rise to the T wave. Note that the T wave is normally the last ECG form in the cardiac cycle; it is followed by the P wave of the next cycle, repeating then the process.

Also of clinical importance in the ECG waveform, intervals like P-R and Q-T are considered for diagnostic purposes. The P-R interval is measured from the beginning of the P wave to the beginning of the QRS complex

and is normally 120-200 ms long. This is basically a measure of the time it takes for an impulse to travel from atrial excitation and through the atria, AV node, and remaining fibers of the conduction system. The Q-T interval is measured from the beginning of the QRS complex to the end of the T wave; this is the time segment from when the ventricles begin their depolarization to the time when they have repolarized to their resting potentials and is normally about 400 ms in duration.

We have now material to bridge the domains of ECG and heart electrophysiology. The interpretation of an ECG involves several subtle details that often exist tacitly in the mind of the cardiologist. Our effort here is to provide a method capable of explicitly uncovering what an ECG maps with respect to canonical heart electrophysiology. We therefore introduce a relation named *maps* meant to associate each of those ECG elementary forms that appears in the ECG to its underlying electrophysiological phenomenon, see Figure 39.

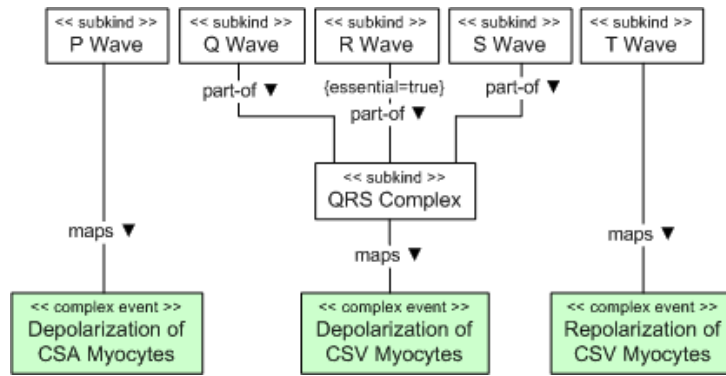


Figure 39: Mapping relations between ECG forms and electrophysiological processes. The colors have no semantic value except to provide the reader with a visual association to the OF-based schemata presented in Section 5.3. The relation *maps* gives a meaning to some of the ECG elementary forms w.r.t. to real electrophysiological phenomena. This model is then of foremost significance in our ECG theory as far as it address the very task of defining an explicit meaning for the ECG.

The *maps* relation can be defined at the instance- and class-level as follows. First, we can formally characterize the relation *observation series of* between an observation series process  $o$  and a (conduction) process  $p$ . The formula below states that if  $o$  is an observation series of process  $p$  then every (atomic) observation which is part of  $o$  is an observation of a part of  $p$  (and can only be an observation of a process which is part of  $p$ ).

$$\text{observation\_series\_of}(o, p) \rightarrow \forall o_1 ( \text{part\_of}(o_1, o) \rightarrow \exists p_1 ( \text{part\_of}(p_1, p) \wedge \text{observation\_of}^{18}(o_1, p_1) ) \\ \wedge \forall p_2 ( \text{observation\_of}(o_1, p_2) \rightarrow \text{part\_of}(p_2, p) ) )$$

In the sequence, we state that if we have two observations  $o_1$  and  $o_2$  which are part of  $o$  and which are observations of parts  $p_1$  and  $p_2$  (parts of  $p$ ), respectively, such that  $o_2$  follows  $o_1$  in the series  $o$  then their respective observed process parts also follow each other in the same way (i.e.,  $p_2$  follows  $p_1$ ).

$$\text{observation\_series\_of}(o, p) \rightarrow \forall o_1, o_2, p_1, p_2 ( \text{part\_of}(o_1, o) \wedge \text{part\_of}(p_1, p) \wedge \text{observation\_of}(o_1, p_1) \\ \wedge \text{part\_of}(o_2, o) \wedge \text{part\_of}(p_2, p) \wedge \text{observation\_of}(o_2, p_2) \wedge \text{follows}(o_2, o_1) \rightarrow \text{follows}(p_2, p_1) )$$

The relation *follows* holding between two processes  $p_2$  and  $p_1$  implies that

$$\text{follows}(p_2, p_1) \rightarrow \exists t_1, t_2 ( \text{last\_instant}(t_1, p_1) \wedge \text{first\_instant}(t_2, p_2) \wedge \text{earlier}(t_1, t_2) )$$

Now, we can characterize the correspondence between an observation series and a sequence of samples

<sup>18</sup>We assume here that if  $\text{observation\_of}(o, p)$  then the process  $o$  occurs either synchronously or after the process  $p$ . Intuitively, there can be no “observation of the future”. This is a simple characterization of a notion (viz., *observation of*) we take actually as somewhat primitive.

representing this series. The first two of these formulae are analogous to formulae just presented for observation series with two important differences. If  $s$  is a sample sequence of observation series  $o$  then: (i) every sample in  $s$  is produced by exactly one observation in  $o$ ; (ii) there is a direct correspondence between observations in  $o$  and samples in  $s$ .

$$\begin{aligned} \text{sample\_sequence\_of}(s, o) &\rightarrow \forall s_1 ( \text{grain\_of}(s_1, s) \rightarrow \exists o_1 ( \text{part\_of}(o_1, o) \wedge \text{produced\_by}(s_1, o_1) ) \wedge \\ &\quad \forall o_2 ( \text{produced\_by}(s_1, o_2) \rightarrow (o_1 = o_2) ) ) \\ \text{sample\_sequence\_of}(s, o) &\rightarrow \forall s_1, s_2, o_1, o_2 ( \text{grain\_of}(s_1, s) \wedge \text{produced\_by}(s_1, o_1) \wedge \text{grain\_of}(s_2, s) \wedge \\ &\quad \text{produced\_by}(s_2, o_2) \wedge \text{successor\_of}(s_2, s_1) \rightarrow \text{directly\_follows}(o_2, o_1) ) \end{aligned}$$

The relation of *successor of* is defined as usual between an element in a sequence and the (direct) successor of that element in that sequence (following the intrinsic ordering criteria of that sequence). The relation of *directly follows* is defined as:

$$\text{directly\_follows}(p_2, p_1) =_{\text{def}} \text{follows}(p_2, p_1) \wedge \neg \exists p_3 ( \text{follows}(p_3, p_1) \wedge \text{follows}(p_2, p_3) )$$

Finally, we can define the relation of *maps* between an elementary form  $c$  and a (conduction) process  $p$ :

$$\text{maps}(c, p) =_{\text{def}} \exists s, o \text{ constituted\_by}(c, s) \wedge \text{sample\_sequence\_of}(s, o) \wedge \text{observation\_series\_of}(o, p)$$

and the corresponding relation at the class-level.

$$\text{maps}(C, P) =_{\text{def}} \forall c ( \text{instance\_of}(c, C) \rightarrow \exists p ( \text{instance\_of}(p, P) \wedge \text{maps}(c, p) ) )$$

## 5.6 From the ECG to Heart Electrophysiology

By employing all the notions just discussed, we have also specified a set of FOL formulae to reconstruct from the ECG waveform the correlated electrophysiological processes occurred over anatomical continuants. These logical assertions make use of our function representations. We start by considering the formulas F1 to F5 given below. They give meaning to the **P wave** based on the function **To conduct CEI** illustrated in the left-hand of Figure 31. So, what are we able to infer once we have a faithfully annotated<sup>19</sup> P wave?

First of all, every **P wave** *maps* one and only one electrophysiological process, viz., the **Depolarization of myocytes of CSA**, see F1. This is just entailed by the model depicted in Figure 39 and the *maps* definition, but still worth to be conveyed here for the sake of clarity.

$$(F1) \forall c \text{ Pwave}(c) \rightarrow \exists p ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$$

Furthermore, every process like this is associated to one and only one CEI and to one and only one CSA playing the role of CEI conductor. Indeed, they need to participate over the whole process. Formally,

$$(F2) \forall p ( \text{DepolarizationOfCSAMyocytes}(p) \rightarrow \exists t_1 \text{ occurring\_at}(p, t_1) \wedge \exists ! c_1, c_2 ( \text{CEI}(c_1) \wedge \\ \forall t ( \text{occurring\_at}(p, t) \rightarrow \text{has\_participant}(p, c_1, t) \wedge \text{CSAasCEIConductor}(c_2, t) \wedge \text{has\_participant}(p, c_2, t) ) ) ) )$$

Nevertheless, if we have the process, we are able to infer that (see F3) there was one state of the world  $SOW_3$  at

<sup>19</sup>Consider a raw ECG waveform, as it is just acquired in some ECG recording session. Often, such an ECG data is further annotated by either a physician or a computer program. Usually, the elementary forms P and T waves and QRS complex are emphasized and even classified - by means of vertical marks on their onset, peak and offset, as well as a (ab)normal signature - to ease pattern-matching in the ECG interpretation. Thus, for “faithfully annotated” we mean such annotations as they are assumed to be trustworthy. The ECG data composing a record is often found annotated in information systems. This is what we are referring to here.

which its requirements have been fulfilled.

$$(F3) \forall p ( \text{DepolarizationOfCSAMyocytes}(p) \rightarrow \\ \exists! c_1, c_2, t_{sow_3} ( \text{CEI}(c_1) \wedge \text{first\_instant}(p, t_{sow_3}) \wedge \text{exists}(c_1, t_{sow_3}) \wedge \text{SANode}(c_2) \wedge \text{located\_in}(c_1, c_2, t_{sow_3}) ) ) )$$

The recognition of the actual realization of To conduct CEI depends on an annotation indicating whether the P wave in hand is normal or not. This can be formally described by F4 as follows. We write “disp\_realized\_by” to denote in fact just *realized\_by*. This convention is used here only for drawing attention to the distinction between a realization (disposition) and its specialization (actual realization). This more specific relation *actually\_realized\_by* is explicitly referred to by means of a binary predicate with the same label.

$$(F4) \forall p, c, f ( ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{ToConductCEI}(f) \wedge \text{disp\_realized\_by}(f, p) \wedge \text{Pwave}(c) \wedge \text{Normal}(c) \wedge \text{maps}(c, p) ) \rightarrow \text{actually\_realized\_by}(f, p) ) )$$

In such case, we can then infer that the goal of To conduct CEI has been fulfilled by the process of Depolarization of myocytes of CSA.

$$(F5) \forall p, f ( ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{ToConductCEI}(f) \wedge \text{actually\_realized\_by}(f, p) ) \\ \rightarrow \exists! c_1, c_2, c_3, t_{sow_4} ( \text{CEI}(c_1) \wedge \text{CSA}(c_2) \wedge \text{conducted\_by}(c_1, c_2) \wedge \text{VentricularPartOfAVBundle}(c_3) \wedge \text{located\_in}(c_1, c_3, t_{sow_4}) \wedge \text{last\_instant}(p, t_{sow_4}) ) ) )$$

Similarly, we can set the formulas F6 - F10 as follows for reconstructing the correlated electrophysiological process from a faithfully annotated QRS complex in the same way. They are based on the right-hand of Figure 31.

$$(F6) \forall c \text{QRScomplex}(c) \rightarrow \exists p ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$$

$$(F7) \forall p ( \text{DepolarizationOfCSVMyocytes}(p) \rightarrow \exists t_1 \text{occurring\_at}(p, t_1) \wedge \exists! c_1, c_2 ( \text{CEI}(c_1) \wedge \\ \forall t ( \text{occurring\_at}(p, t) \rightarrow \text{has\_participant}(p, c_1, t) \wedge \text{CSVasCEIConductor}(c_2, t) \wedge \text{has\_participant}(p, c_2, t) ) ) )$$

$$(F8) \forall p ( \text{DepolarizationOfCSVMyocytes}(p) \rightarrow \exists! c_1, c_2, t_{sow_5} ( \text{CEI}(c_1) \wedge \text{first\_instant}(p, t_{sow_5}) \wedge \text{exists}(c_1, t_{sow_5}) \\ \wedge \text{VentricularPartOfAVBundle}(c_2) \wedge \text{located\_in}(c_1, c_2, t_{sow_5}) ) ) )$$

$$(F9) \forall p, c, f ( ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{ToConductCEI}(f) \wedge \text{disp\_realized\_by}(f, p) \\ \wedge \text{QRScomplex}(c) \wedge \text{Normal}(c) \wedge \text{maps}(c, p) ) \rightarrow \text{actually\_realized\_by}(f, p) ) )$$

$$(F10) \forall p, f ( ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{ToConductCEI}(f) \wedge \text{actually\_realized\_by}(f, p) ) \\ \rightarrow \exists c_1, c_2 ( \text{CEI}(c_1) \wedge \text{CSV}(c_2) \wedge \text{conducted\_by}(c_1, c_2) ) ) )$$

However, an additional formula F11 also holds for representing a relation between the atrial and ventricular manifestations of the function To conduct CEI, cf. Figure 31. If an instance of this function has been actually realized by an individual of CSA and other individual of CSV, then it follows that a CEI individual - which has been *located in* the Ventricular part of AV bundle at  $SOW_4$  and at  $SOW_5$  - these two states of affairs are then identical - has been *conducted by* both of the CSV and CSA individuals. That is, the CEI conducted by CSV is not originated from a escape beat.

$$(F11) \forall f, p_1, p_2 ( ( \text{ToConductCEI}(f) \wedge \text{DepolarizationOfCSAMyocytes}(p_1) \wedge \text{actually\_realized\_by}(f, p_1) \\ \wedge \text{DepolarizationOfCSVMyocytes}(p_2) \wedge \text{actually\_realized\_by}(f, p_2) ) \\ \rightarrow \exists c, c_1, c_2 ( \text{CEI}(c) \wedge \text{CSA}(c_1) \wedge \text{conducted\_by}(c, c_1) \wedge \text{CSV}(c_2) \wedge \text{conducted\_by}(c, c_2) ) ) )$$

Another germane relation between functions holds between the atrial manifestation of To conduct CEI and To generate CEI. We can establish it by relying on a bit indirect account from the P wave. Even then the

pacemaker SA node myocytes' excitation is not visible in the ECG waveform, the presence of the P wave provides, if not a proof, a strong evidence for the actual realization of To generate CEI by the SA node. Indeed, the most significative connection to be established is that the CEI *located in* the SA node at  $SOW_3$  satisfying the requirement of To conduct CEI around the atria is actually the same *generated by* the SA node as CEI generator and *located in* the SA node at  $SOW_2$  (the goal of function To generate CEI), i.e.,  $SOW_3$  and  $SOW_2$  actually coincide.

$$(F12) \forall p, t_{sow_3}, c_1, c_2 ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{first\_instant}(p, t_{sow_3}) \wedge \text{CEI}(c_1) \wedge \text{exists}(c_1, t_{sow_3}) \wedge \text{SANode}(c_2) \\ \wedge \text{located\_in}(c_1, c_2, t_{sow_3}) \rightarrow \exists! c_3, p_1, t_{sow_2} ( \text{SANode}(c_3) \wedge \text{generated\_by}(c_1, c_3) \\ \wedge \text{DepolarizationOfPacemakerSANodeMyocytes}(p_1) \wedge \text{last\_instant}(p_1, t_{sow_2}) \wedge \text{located\_in}(c_1, c_2, t_{sow_2}) ) ) )$$

We can then set the formulas F13 - F15 to explicitly assert the consequences of that. First, (F13) the accomplishment of the goal of To generate CEI entails it has been actually realized.

$$(F13) \forall c, c_1, c_2, p, t_{sow_2} ( ( \text{CEI}(c) \wedge \text{SANode}(c_1) \wedge \text{generated\_by}(c, c_1) \\ \wedge \text{DepolarizationOfPacemakerSANodeMyocytes}(p) \wedge \text{last\_instant}(p, t_{sow_2}) \wedge \text{SANode}(c_2) \wedge \text{located\_in}(c, c_2, t_{sow_2}) ) \\ \rightarrow \exists f ( \text{ToGenerateCEI}(f, p) \wedge \text{actually\_realized\_by}(f, p) ) ) )$$

But also, it holds as well the assertion of the usual formulas for any function we have represented: (F14) the connection between the process of Depolarization of Pacemaker SA node myocytes and its participants; (F15) the inference that if we have that process then there was one state of the world  $SOW_1$  at which its requirements have been fulfilled.

$$(F14) \forall p ( \text{DepolarizationOfPacemakerSANodeMyocytes}(p) \rightarrow \exists t_1 \text{ occurring\_at}(p, t_1) \wedge \exists! c_2 ( \text{SANode}(c_2) \wedge \\ \forall t ( \text{occurring\_at}(p, t) \rightarrow \text{SANodeAsCEIGenerator}(c_2, t) \wedge \text{has\_participant}(p, c_2, t) ) ) \wedge \exists! c_1 \exists t_{sow_2} ( \text{CEI}(c_1) \\ \wedge \text{last\_instant}(p, t_{sow_2}) \wedge \text{has\_participant}(p, c_1, t_{sow_2}) \wedge \forall t' \text{ has\_participant}(p, c_1, t') \rightarrow (t' = t_{sow_2}) ) ) )$$

$$(F15) \forall p ( \text{DepolarizationOfPacemakerSANodeMyocytes}(p) \rightarrow \exists! t_{sow_1}, c_1, \neg \exists c_2 ( \text{first\_instant}(p, t_{sow_1}) \wedge \\ \text{PacemakerSANodeMyocytesPolarized}(c_1, t_{sow_1}) \wedge \text{CEI}(c_2, t_{sow_1}) ) ) )$$

Finally, the electrophysiological process of Repolarization of myocytes of CSV is (loosely speaking) mapped by the T wave in the ECG waveform. This process is a realization of the function To restore EPs. Thus, once a T wave has been recognized, by the same token we can infer the following as stated by formulas F16 - F20.

$$(F16) \forall c \text{ Twave}(c) \rightarrow \exists p ( \text{RepolarizationOfCSVMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$$

$$(F17) \forall p ( \text{RepolarizationOfCSVMyocytes}(p) \rightarrow \exists t_1 \text{ occurring\_at}(p, t_1) \wedge \exists! c ( \text{CSV}(c) \wedge \\ \forall t ( \text{occurring\_at}(p, t) \rightarrow \text{CSVasEPsAccumulator}(c, t) \wedge \text{has\_participant}(p, c, t) ) ) )$$

$$(F18) \forall p ( \text{RepolarizationOfCSVMyocytes}(p) \rightarrow \exists! t_{sow_7}, c ( \text{first\_instant}(p, t_{sow_7}) \wedge \text{CSVMyocytesDepolarized}(c, t_{sow_7}) ) )$$

$$(F19) \forall p, f, c ( ( \text{RepolarizationOfCSVMyocytes}(p) \wedge \text{ToRestoreEPs}(f) \wedge \text{disp\_realized\_by}(f, p) \\ \wedge \text{Twave}(c) \wedge \text{Normal}(c) \wedge \text{maps}(c, p) ) \rightarrow \text{actually\_realized\_by}(f, p) ) )$$

$$(F20) \forall p, f ( ( \text{RepolarizationOfCSVMyocytes}(p) \wedge \text{ToRestoreEPs}(f) \wedge \text{actually\_realized\_by}(f, p) ) \\ \rightarrow \exists! t_{sow_8}, c ( \text{last\_instant}(p, t_{sow_8}) \wedge \text{CSVMyocytesPolarized}(c, t_{sow_8}) ) ) )$$

The formulae just presented can be implemented and then used for automated reasoning over universals and particulars of the ECG theory developed hitherto. This is in fact the business of a reasoning-based application we have developed which is presented in Chapter 8.



## 5.7 An ECG Ontology

The results of our ontological study of the electrocardiogram have been the source of domain knowledge in the construction of the **ECG Ontology**. It constitutes a solution-independent theory of the ECG, which is meant to be reused across multiple applications. The ECG Ontology handles what the ECG is on both sides of the patient and of the physician. As we have seen, that relies on a number of notions related to the heart electrophysiology, which takes place over anatomical entities. The ECG Ontology then comes together with two extra original sub-ontologies, viz., the anatomy for ECG and heart electrophysiology sub-ontologies. It also imports the OBO Relation Ontology (RO), which is then extended by the relations defined in our theory, see Figure 40.

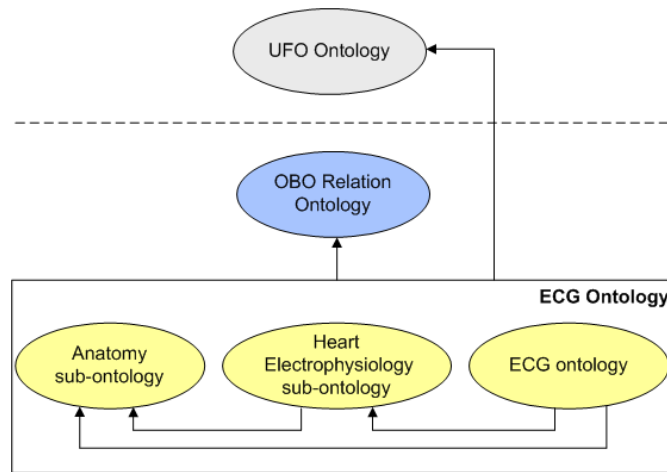


Figure 40: Import relationships of the ECG Ontology. The arrows point towards the ontology being imported. The UFO ontology is used to ground the ECG domain entities in a sound ontological basis. The OBO Relation Ontology is imported here to provide us with basic relations as they are standardized in the biomedical domain.

### 5.7.1 Competence Questions

The scope of the ECG Ontology can be defined by means of the following competence questions (CQ).

- CQ1.** What essentially composes an ECG record?
- CQ2.** What is the source of an ECG record, i.e., how is it obtained?
- CQ3.** What is object of physician's analysis in the ECG waveform for interpreting a correlated heart behavior?
- CQ4.** For all ECG elementary forms, which heart electrophysiological process(es) does (do) it map (if at all)?
- CQ5.** For all heart electrophysiological functions, which anatomical entity(ies) is (are) able to realize it?
- CQ6.** For all heart electrophysiological functions, which requirements must be satisfied to enable its realization?
- CQ7.** For all heart electrophysiological functions, which goals must be satisfied to accomplish its realization?

As suggested by Uschold and Gruninger (62), CQs provide conditions for evaluating the ontology effectiveness and completeness. Moreover, they prescribe such an evaluation process to be carried out in a formal fashion as long as the CQs are stated in formal logic. Along these lines, the ECG Ontology CQs above have been formally

described in FOL as follows. These competence questions then delimitate our domain in an objective way.

- CQ1.**  $\forall c ( \text{Record}(c) \rightarrow \exists c_1 ( \text{Waveform}(c_1) \wedge \text{essential\_part\_of}(c_1, c) ) )$
- CQ2.**  $\forall c ( \text{Record}(c) \rightarrow \exists p, c_1, c_2 ( \text{RecordingSession}(p) \wedge \text{produced\_by}(c, p) \wedge \text{RecordingDevice}(c_1) \wedge \text{Person}(c_2) \wedge \forall t \text{ occurring\_at}(p, t) \rightarrow ( \text{RDAsRecorder}(c_1, t) \wedge \text{has\_participant}(p, c_1) \wedge \text{Patient}(c_2, t) \wedge \text{has\_participant}(p, c_2) ) ) )$
- $\forall c ( \text{Record}(c) \rightarrow \exists w, s, p ( \text{Waveform}(w) \wedge \text{essential\_part\_of}(w, c) \wedge \text{SampleSequence}(s) \wedge \text{constituted\_by}(w, s) \wedge \text{ObservationSeries}(p) \wedge \text{sample\_sequence\_of}(s, p) \wedge \forall p_1, t_1 ( ( \text{Observation}(p_1) \wedge \text{occurring\_at}(p_1, t_1) \wedge \text{part\_of}(p_1, p) ) \rightarrow \exists e, l, bs, hb, pa, rs, rd ( \text{ElectrodeAsMeasurer}(e, t_1) \wedge \text{has\_participant}(p_1, e) \wedge \text{Lead}(l) \wedge \text{has\_participant}(p_1, l) \wedge \text{BodySurfaceRegionAsObjectOfMeasure}(bs, t_1) \wedge \text{has\_participant}(p_1, bs) \wedge \text{part\_of}(bs, hb) \wedge \text{Patient}(pa, t_1) \wedge \text{constitutes}(hb, pa) \wedge \text{RecordingSession}(rs) \wedge \text{participates}(pa, rs) \wedge \text{RecordingDevice}(rd) \wedge \text{part\_of}(e, rd) \wedge \text{participates}(rd, rs) \wedge \text{produces}(rs, c) ) ) ) )$
- CQ3.**  $\forall c \text{ ElementaryForm}(c) \rightarrow ( \text{Cycle}(c) \vee \text{Wave}(c) \vee \text{Segment}(c) \vee \text{QRScomplex}(c) \vee \text{Baseline}(c) )$
- $\forall c \text{ Wave}(c) \rightarrow ( \text{Pwave}(c) \vee \text{Qwave}(c) \vee \text{Rwave}(c) \vee \text{Swave}(c) \vee \text{Twave}(c) )^{20}$
- $\forall c \text{ Segment}(c) \rightarrow ( \text{PRsegment}(c) \vee \text{STsegment}(c) \vee \text{TPsegment}(c) )$
- CQ4.**  $\forall c \text{ Pwave}(c) \rightarrow \exists p ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$  (R1)
- $\forall c \text{ QRScomplex}(c) \rightarrow \exists p ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$  (R7)
- $\forall c \text{ Twave}(c) \rightarrow \exists p ( \text{RepolarizationOfCSVMyocytes}(p) \wedge \text{maps}(c, p) \wedge \forall p_1 ( \text{maps}(c, p_1) \rightarrow (p_1 = p) ) )$  (R19)
- CQ5.**  $\forall f ( \text{ToGenerateCEI}(f) \rightarrow \exists c_1 ( \text{SANode}(c_1) \wedge \text{characterized\_by}(c_1, f) ) ) \wedge \forall c ( \text{SANode}(c) \rightarrow \exists f_1 ( \text{ToGenerateCEI}(f_1) \wedge \text{characterized\_by}(c, f_1) ) )$
- $\forall f ( \text{ToConductCEI}(f) \rightarrow \exists c_1, c_2 ( \text{CSA}(c_1) \wedge \text{characterized\_by}(c_1, f) \wedge \text{CSV}(c_2) \wedge \text{characterized\_by}(c_2, f) ) ) \wedge \forall c_1, c_2 ( \text{CSA}(c_1) \wedge \text{CSV}(c_2) \rightarrow \exists f_1 ( \text{ToConductCEI}(f_1) \wedge \text{characterized\_by}(c_1, f_1) \wedge \text{characterized\_by}(c_2, f_1) ) )$
- $\forall f ( \text{ToRestoreEPs}(f) \rightarrow \exists c_1 ( \text{CSV}(c_1) \wedge \text{characterized\_by}(c_1, f) ) ) \wedge \forall c ( \text{CSV}(c) \rightarrow \exists f_1 ( \text{ToGenerateCEI}(f_1) \wedge \text{characterized\_by}(c, f_1) ) )$
- CQ6.**  $\forall f ( \text{ToGenerateCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfPacemakerSANodeMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \exists t_{sow_1} ( \text{first\_instant}(p, t_{sow_1}) \rightarrow \exists c_1 \neg \exists c_2 ( \text{PacemakerSANodeMyocytesPolarized}(c_1, t_{sow_1}) \wedge \text{CEI}(c_2, t_{sow_1}) ) ) ) )$
- $\forall f ( \text{ToConductCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \exists t_{sow_3} ( \text{first\_instant}(p, t_{sow_3}) \rightarrow \exists c_1, c_2 ( \text{CEI}(c_1) \wedge \text{exists}(c_1, t_{sow_3}) \wedge \text{SANode}(c_2) \wedge \text{located\_in}(c_1, c_2, t_{sow_3}) ) ) ) )$
- $\forall f ( \text{ToConductCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \exists t_{sow_5} ( \text{first\_instant}(p, t_{sow_5}) \rightarrow \exists c_1, c_2 ( \text{CEI}(c_1) \wedge \text{exists}(c_1, t_{sow_5}) \wedge \text{VentricularPartOfAVBundle}(c_2) \wedge \text{located\_in}(c_1, c_2, t_{sow_5}) ) ) ) )$
- $\forall f ( \text{ToRestoreEPs}(f) \rightarrow \exists p ( \text{RepolarizationOfCSVMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \exists t_{sow_7} ( \text{first\_instant}(p, t_{sow_7}) \rightarrow \exists c_1 \text{PacemakerSANodeMyocytesDepolarized}(c_1, t_{sow_7}) ) ) )$
- CQ7.**  $\forall f ( \text{ToGenerateCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfPacemakerSANodeMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \text{actually\_realized\_by}(f, p) \leftrightarrow \exists t_{sow_2}, c_1, c_2 ( \text{last\_instant}(p, t_{sow_2}) \wedge \text{CEI}(c_1) \wedge \text{SANode}(c_2) \wedge \text{generated\_by}(c_1, c_2) \wedge \text{located\_in}(c_1, c_2, t_{sow_2}) ) ) )$
- $\forall f ( \text{ToConductCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfCSAMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \text{actually\_realized\_by}(f, p) \leftrightarrow \exists t_{sow_4}, c_1, c_2, c_3 ( \text{last\_instant}(p, t_{sow_4}) \wedge \text{CEI}(c_1) \wedge \text{CSA}(c_2) \wedge \text{conducted\_by}(c_1, c_2) \wedge \text{VentricularPartOfAVBundle}(c_3) \wedge \text{located\_in}(c_1, c_3, t_{sow_4}) ) ) )$
- $\forall f ( \text{ToConductCEI}(f) \rightarrow \exists p ( \text{DepolarizationOfCSVMyocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \wedge \text{actually\_realized\_by}(f, p) \leftrightarrow \exists t_{sow_6}, c_1, c_2 ( \text{last\_instant}(p, t_{sow_6}) \wedge \text{CEI}(c_1) \wedge \text{CSV}(c_2) \wedge \text{conducted\_by}(c_1, c_2) ) ) )$

<sup>20</sup>We are assuming not to exist a U wave with relevance for physician's analysis.

$$\begin{aligned} & \forall f ( \text{ToRestoreEPs}(f) \rightarrow \exists p ( \text{RepolarizationOfCSVMycocytes}(p) \wedge \text{disp\_realized\_by}(f, p) \\ & \wedge \text{actually\_realized\_by}(f, p) \leftrightarrow \exists t_{\text{sowg}}, c_1 ( \text{last\_instant}(p, t_{\text{sowg}}) \wedge \text{CSVMycocytesPolarized}(c_1, t_{\text{sowg}}) ) ) ) \end{aligned}$$

By relying on the machinery achieved with the ECG Ontology implementation presented in Chapter 6, we address an ontology evaluation which also considers these formal CQs to verify the ontology competence.

## 5.7.2 Documentation

In addition to the models and FOL formulae presented hitherto in this chapter, we provide a documentation comprising: (i) all relations coined here by us or taken from sources other than the OBO RO, and their meta-properties; we have these relations as an extension to the OBO RO ontology required for the ECG Ontology, except by some of them which are specific to the ECG domain (viz., *observation of*, *observation series of*, *sample sequence of* and *maps*); (ii) a class dictionary with corresponding term, class ID, UFO type and textual definition for the ECG Ontology entities. An updated status of the ECG Ontology is available at our project website<sup>21</sup>.

The relations are presented in Table 2. With the purpose of easing comprehensibility, we divide the class dictionary into several parts according to the corresponding sub-ontologies they relate to. We start with the anatomy sub-ontology classes, which are gathered into Table 3, and proceed to catalogue the heart electrophysiology classes in Table 4. The classes relating to the ECG ontology are then presented in Table 5. Notice, especially in the two latter tables, that the class textual definitions often point to other classes in an attempt to rule out redundancy.

Table 3: ECG Ontology class dictionary: sub-ontology of anatomy. Relations and classes of the ECG Ontology appear in the textual definitions emphasized.

Term	Class ID	UFO Type	Textual Definition
Anatomical entity	ecgOnto:001	category	From FMAID:62955. Most general continuant entity in the anatomy sub-ontology, which subsumes all other entities. Examples: organ component, human body, heart, portion of tissue, SA node.
Material anatomical entity	ecgOnto:002	category	From FMAID:67165. Subsumes all concrete <i>anatomical entities</i> , in the sense of Aristotle's <i>substance</i> , or matter. Examples: organ component, human body, heart, portion of tissue, SA node.
Immaterial anatomical entity	ecgOnto:003	category	Similar to FMAID:67112. Subsumes all physical <i>anatomical entities</i> which are three-dimensional space, surface, line or point existentially dependent on some <i>material anatomical entity</i> . Examples: body space, surface of heart, costal margin, apex of right lung, anterior compartment of right arm.
Anatomical boundary entity	ecgOnto:004	category	Similar to FMAID:50705. <i>Immaterial anatomical entity</i> of one less dimension than the <i>anatomical entity</i> it bounds or demarcates from another <i>anatomical entity</i> . Examples: surface of heart, surface of epithelial cell, cervicothoracic plane, supra-orbital notch, costal margin, apex beat, Sylvian point.
Body surface	ecgOnto:005	kind	Similar FMAID:61695. Bona fide <i>anatomical boundary entity</i> , which is the external surface of the whole body. Examples: There is only one body surface.

Continued on next page

<sup>21</sup><http://nemo.inf.ufes.br/biomedicine/ecg.html>.

Table 3 – continued from previous page

Term	Class ID	UFO Type	Textual Definition
Body surface region	ecgOnto:006	kind	Similar to FMAID:24146. A fiat region <i>part of</i> the <i>body surface</i> . It is the external surface of a body part or a body part subdivision. Examples: surface of head, surface of front of neck, epigastric region, sacral region, surface of dorsum of right foot.
Human body	ecgOnto:007	kind	From FMAID:20394. <i>Material anatomical entity</i> which an individual member of the human species is <i>constituted by</i> . Examples: There is only one human body.
Organ	ecgOnto:008	category	From FMAID:67498. <i>Material anatomical entity</i> which <i>has as parts</i> portions of two or more types of tissue or two or more types of cardinal organ part which constitute a maximally connected anatomical structure demarcated predominantly by a bona fide anatomical surface. Examples: femur, biceps, liver, heart, skin, tracheobronchial tree, ovary.
Organ system	ecgOnto:009	category	From FMAID:7149. <i>Material anatomical entity</i> that <i>has as parts</i> one or more organ types which are interconnected with one another by zones of continuity. Examples: skeletal system, cardiovascular system, alimentary system.
Organ component	ecgOnto:010	category	From FMAID:14065. <i>Material anatomical entity</i> that is <i>part of</i> an <i>organ</i> and bounded predominantly by bona fide <i>anatomical boundary entities</i> . Examples: lobe of lung, osteon, acinus, submucosa, anterior leaflet of mitral valve, capsule of kidney, cortical bone, muscle fasciculus.
Region of organ component	ecgOnto:011	category	From FMAID:86103. <i>Material anatomical entity</i> that is a fiat region <i>part of</i> an <i>organ component</i> . Examples: cervical part of wall of esophagus, mucosa of body of stomach.
Portion of tissue	ecgOnto:012	category	Similar to FMAID:9637. <i>Material anatomical entity constituted by</i> two types of <i>portions of body substance</i> , viz., cells and extracellular matrix. Examples: epithelium, muscle tissue, connective tissue, neural tissue, lymphoid tissue.
Heart	ecgOnto:013	kind	Inspired in FMAID:7088. <i>Organ</i> with cavitated <i>organ components</i> (it <i>has as parts organ chambers</i> , which is continuous with the systemic and pulmonary arterial and venous trees. Examples: There is only one heart.
Cardiovascular system	ecgOnto:014	kind	From FMAID:7161. <i>Organ system</i> that <i>has as parts</i> the <i>heart</i> , the systemic and pulmonary arterial and venous system, the lymphatic and the portal venous system.
Wall of organ	ecgOnto:015	category	From FMAID:82482. <i>Organ component</i> adjacent to an organ cavity and which consists of a maximal aggregate of organ component layers.
Organ chamber	ecgOnto:016	category	From FMAID:82481. Cavitated type of <i>organ component</i> .
Muscle layer of organ	ecgOnto:017	category	From FMAID:85353. Layer of organ constituted by a muscle tissue.
Wall of heart	ecgOnto:018	kind	From FMAID:7274. <i>Wall of organ</i> which <i>has as its parts</i> the endocardium, <i>myocardium</i> , epicardium, and also the cardiac septum.

Continued on next page

Table 3 – continued from previous page

Term	Class ID	UFO Type	Textual Definition
Right atrium	ecgOnto:019	kind	From FMAID:7096. Right (cardiac) <i>organ chamber</i> which is continuous with the superior vena cava and inferior vena cava.
Right ventricle	ecgOnto:020	kind	From FMAID:7098. Right (cardiac) <i>organ chamber</i> which is continuous with the pulmonary arterial trunk.
Left atrium	ecgOnto:021	kind	From FMAID:7097. Left (cardiac) <i>organ chamber</i> which is continuous with the pulmonary venous trunk.
Left ventricle	ecgOnto:022	kind	From FMAID:7101. Left (cardiac) <i>organ chamber</i> which is continuous with the aorta.
Myocardium	ecgOnto:023	kind	From FMAID:9462. <i>Muscle layer of organ</i> which has as part the heart conducting system.
Region of wall of heart	ecgOnto:024	category	Similar to FMAID:86212. Fiat region of the <i>wall of heart</i> that bounds some (cardiac) <i>organ chamber</i> .
Region of myocardium	ecgOnto:025	category	Similar to FMAID:86044. Fiat region of the <i>myocardium part</i> of some <i>region of wall of heart</i> .
Wall of right atrium	ecgOnto:026	kind	From FMAID:9457. Atrial <i>region of wall of heart</i> which is continuous with the wall of superior vena cava an inferior vena cava.
Wall of left atrium	ecgOnto:027	kind	From FMAID:9531. Atrial <i>region of wall of heart</i> which is continuous with the wall of pulmonary vein.
Wall of left ventricle	ecgOnto:028	kind	From FMAID:9556. Ventricular <i>region of wall of heart</i> which is continuous with the wall of aorta.
Wall of right ventricle	ecgOnto:029	kind	From FMAID:9533. Ventricular <i>region of wall of heart</i> which is continuous with the wall of pulmonary trunk.
Right atrial myocardium	ecgOnto:030	kind	From FMAID:83531. Atrial <i>region of myocardium</i> that has as part the <i>conducting system of the right atrium</i> and is is continuous with the tunica media of superior and inferior vena cavae.
Left atrial myocardium	ecgOnto:031	kind	From FMAID:83532. Atrial <i>region of myocardium</i> that has as part the <i>left branch of Bachmann's bundle</i> and is continuous with the tunica media of pulmonary vein.
Left ventricular myocardium	ecgOnto:032	kind	From FMAID:9558. Ventricular <i>region of myocardium</i> that has as part the <i>left bundle branch</i> and is continuous with the tunica media of aorta.
Right ventricular myocardium	ecgOnto:033	kind	From FMAID:9535. Ventricular <i>region of myocardium</i> that has as part the <i>right bundle branch</i> and is continuous with the tunica media of the pulmonary trunk.
Conducting system of heart	ecgOnto:034	kind	From FMAID:9476. Conducting tissue of heart which is <i>constituted</i> by extracellular matrix and specialized cardiac myocytes in the <i>myocardium</i> .
Conducting system of subdivision of heart	ecgOnto:035	category	From FMAID:83513. Conducting system of a fiat region of the <i>heart</i> , viz., the <i>CSA</i> , <i>CSV</i> and the <i>conducting system of the right atrium</i> .
Subdivision of conducting system of heart	ecgOnto:036	category	From FMAID:6266. Fiat subdivision of the <i>conducting system of heart</i> .

Continued on next page

Table 3 – continued from previous page

Term	Class ID	UFO Type	Textual Definition
Conducting system of right atrium	ecgOnto:037	kind	Similar to FMAID:13877. Conducting system of subdivision of the <i>heart</i> that is part of the <i>CSA</i> . It is the mereological sum of the <i>SA node</i> , <i>AV node</i> , <i>atrial part of AV bundle</i> , <i>right branch of Bachmann's bundle</i> , as well as the <i>posterior</i> , <i>middle</i> and <i>anterior</i> internodal tracts.
Conducting system of atria	ecgOnto:038	kind	<i>CSA</i> . Mereological sum of the <i>conducting system of the right atrium</i> and the conducting system of the left atrium (FMAID:13878). The latter coincides in the ECG Ont. with the <i>left branch of Bachmann's bundle</i> .
Conducting system of ventricles	ecgOnto:039	kind	<i>CSV</i> . Mereological sum of the conducting system of the right ventricle (FMAID:13879) and that of the left ventricle (FMAID:13880).
Atrial part of AV bundle	ecgOnto:040	kind	From FMAID:9540. Fiat <i>subdivision of the conducting system of the heart</i> which enters the central fibrous body and is continuous with the <i>ventricular part of AV bundle</i> (the AV bundle is also referred to as bundle of His).
Right branch of Bachmann's bundle	ecgOnto:041	kind	Fiat part of internodal tract (a <i>subdivision of the conducting system of the heart</i> ) which is similar to the (FMAID:83346) atrial septal branch of anterior internodal tract; but only touches on the <i>anterior tract</i> (it does not intersects with it). It is continuous with the <i>AV node</i> .
Posterior tract	ecgOnto:042	kind	From FMAID:9483. Also known as Thorel pathway. Internodal tract (a <i>subdivision of the conducting system of the heart</i> ) that provides a conduction pathway from the <i>SA node</i> to the <i>AV node</i> by crossing the <i>right atrial myocardium</i> roughly through the right part of the <i>wall of heart</i> .
Middle tract	ecgOnto:043	kind	From FMAID:9482. Also known as Wenckebach pathway. Internodal tract (a <i>subdivision of the conducting system of the heart</i> ) that provides a conduction pathway from the <i>SA node</i> to the <i>AV node</i> by crossing the <i>right atrial myocardium</i> roughly through its center.
Anterior tract	ecgOnto:044	kind	From FMAID:9480. Internodal tract (a <i>subdivision of the conducting system of the heart</i> ) that extends from the anterior part of the <i>SA node</i> and descends along the <i>right atrium</i> to connect to the <i>AV node</i> . It touches the right branch of Bachmann's bundle.
SA node	ecgOnto:045	kind	Inspired in FMAID:9477. Subdivision of the <i>conducting system of heart</i> located at the junction of the right atrium and the superior vena cava ("the roof of the RA"), around the sinoatrial nodal branch of right coronary artery and is continuous with the internodal tract.
AV node	ecgOnto:046	kind	Inspired in FMAID:9478. Subdivision of the <i>conducting system of heart</i> which is located in the muscular part of the interatrial septum that is continuous with the <i>AV bundle</i> .
Left branch of Bachmann's bundle	ecgOnto:047	kind	Internodal tract that starts in the <i>right atrial myocardium</i> from a bifurcation at the anterior tract and terminates in the <i>left atrial myocardium</i> . It is similar to the atrial branch of anterior internodal tract (FMAID:84578).

Continued on next page

Table 3 – continued from previous page

Term	Class ID	UFO Type	Textual Definition
Ventricular part of AV bundle	ecgOnto:048	kind	Similar to FMAID:9541. Subdivision of the <i>conducting system of heart</i> which is located in the muscular part of the interventricular septum and is continuous with both the <i>right bundle branch</i> and the <i>left bundle branch</i> .
Left bundle branch	ecgOnto:049	kind	<i>Subdivision of the conducting system of heart</i> which enters the septal <i>left ventricular myocardium</i> and is continuous with the anterior, septal and posterior divisions of the <i>left bundle branch</i> . Similar to left branch of atrioventricular bundle (FMAID:9487).
Right bundle branch	ecgOnto:050	kind	<i>Subdivision of conducting system of heart</i> which enters the septal myocardium of <i>right ventricle</i> to reach the anterior papillary muscle and then subendocardially to the apex of the heart. Similar to right branch of atrioventricular bundle (FMAID:9486).
Anatomical cluster	ecgOnto:051	category	Similar to FMAID:49443. Entity which anatomical structures emerge from. It <i>has as parts</i> a heterogeneous collective of organs, organ parts, cells, cell parts or body part subdivisions that are adjacent to, or continuous with one another. Examples: cells of the heart, joint, adnexa of uterus, root of lung, renal pedicle.
Cell cluster	ecgOnto:052	category	Similar to FMAID:62807. <i>Anatomical cluster</i> which <i>has as grains</i> many cells grouped together according to shared attributes.
SA node myocytes	ecgOnto:053	collective	<i>Cell cluster</i> that <i>partially constitutes</i> the <i>SA node</i> and is a <i>subcollection</i> of the <i>CSA myocytes</i> .
Pacemaker SA node myocytes	ecgOnto:054	collective	<i>Cell cluster</i> that is a <i>subcollection</i> of the <i>SA node myocytes</i> whose cells bear the primary property of spontaneously depolarizing.
Transitional SA node myocytes	ecgOnto:055	collective	<i>Cell cluster</i> that is a <i>subcollection</i> of the <i>SA node myocytes</i> whose cells bear the primary property propagating action potentials (or the <i>cardiac electrical impulse</i> proper).
CSA Myocytes	ecgOnto:056	collective	<i>Cell cluster</i> that <i>partially constitutes</i> the <i>CSA</i> .
CSV Myocytes	ecgOnto:057	collective	<i>Cell cluster</i> that <i>partially constitutes</i> the <i>CSV</i> .
Portion of body substance	ecgOnto:058	category	From FMAID:9669. <i>Material anatomical entity</i> in a gaseous, liquid, semisolid or solid state, with or without the admixture of cells and biological macromolecules; produced by anatomical structures or derived from inhaled and ingested substances that have been modified by anatomical structures. Examples: ECM, saliva, semen, cerebrospinal fluid, respiratory air, urine, feces, blood, plasma, lymph.
Portion of extracellular matrix	ecgOnto:059	category	From FMAID:9672. Body substance that is a fluid matrix where cells are immersed in. It consists of ground substance and connective tissue fibers.
ECM of SA node	ecgOnto:060	quantity	<i>Portion of extracellular matrix</i> that <i>partially constitutes</i> the <i>SA node</i> .
ECM of CSA	ecgOnto:061	quantity	<i>Portion of extracellular matrix</i> that <i>partially constitutes</i> the <i>CSA</i> .
ECM of CSV	ecgOnto:062	quantity	<i>Portion of extracellular matrix</i> that <i>partially constitutes</i> the <i>CSV</i> .

Table 2: ECG Ontology relations and their meta-properties.

<i>Relation</i>	<i>Reflexivity</i>	<i>Asymmetry</i>	<i>Transitivity</i>	<i>Inverse of</i>	<i>Description</i>
<i>part_of</i>	-	+	+	<i>has_part</i>	Proper parthood relation between two continuants.
<i>subcollection_of</i>	-	+	+	<i>has_subcollection</i>	Type of proper parthood holding between two collective entities. It holds also the weak supplementation property.
<i>grain_of</i>	-	+	-	<i>has_grain</i>	Unusual type of parthood to capture the notion of a continuant at a given level of granularity that is a member of a collective lied at the next granularity level. It holds also the weak supplementation property.
<i>constituted_by</i>	-	+	+	<i>constitutes</i>	Relation between two continuants that denotes to some extent the notion of emergence. Constitution is not identity. Source: DOLCE (48).
<i>partially_constituted_by</i>	-	+	+	<i>partially_constitutes</i>	Specific type of constitution where at least two entities are necessary to a third one to emerge.
<i>characterized by</i>	-	+	-	<i>characterizes</i>	Class-level relation for the instance-level relation of <i>inherence</i> . Characterization holds between a substantial continuant and another continuant that inheres in it. Consequently, this relation stands for the existential dependence property.
<i>realized_by</i>	-	+	-	<i>realizes</i>	Relation between a function and a process which bears the disposition to realize it.
<i>actually_realized_by</i>	-	+	-	-	Specializes <i>realized_by</i> as the function in the domain has been in fact realized by the process in the range.
<i>produced_by</i>	-	+	-	-	Relation between a continuant and a process in which the continuant participates only at the process' end time instant. That is, the continuant has been produced by the process.
<i>generated_by</i>	-	+	-	-	Relation between two continuants such that the continuant in the range participates as an agent in some process which the continuant in the domain is <i>produced by</i> .
<i>conducted_by</i>	-	+	-	-	Relation between two continuants such that the continuant in the domain is an instance of <i>mode</i> and that in the range is an instance of <i>kind</i> .
<i>sample_sequence_of</i>	-	+	-	-	Relation between a <b>Sample sequence</b> and an <b>Observation series</b> . Every individual of the former is <i>sample sequence of</i> some individual of the latter.
<i>mediates</i>	-	+	-	-	Relation between a relator entity and some entity that must have some qua individual inhering in it.
<i>maps</i>	-	+	-	-	Relation between an ECG elementary form and an electrophysiological process which can be apprehended by it.



Table 4: ECG Ontology class dictionary: sub-ontology of Heart Electrophysiology. Relations and classes of the ECG Ontology appear in the textual definitions emphasized.

<i>Term</i>	<i>Class ID</i>	<i>UFO type</i>	<i>Textual Definition</i>
SA node myocytes polarized	ecgOnto:063	phase	Phase of the <i>SA node myocytes</i> where they are in a polarized, or resting state.
SA node myocytes depolarized	ecgOnto:064	phase	Phase of the <i>SA node myocytes</i> where they are in a depolarized state.
CSA myocytes polarized	ecgOnto:065	phase	Phase of the <i>CSA myocytes</i> where they are in a polarized, or resting state.
CSA myocytes depolarized	ecgOnto:066	phase	Phase of the <i>CSA myocytes</i> where they are in a depolarized state.
CSV myocytes polarized	ecgOnto:067	phase	Phase of the <i>CSV myocytes</i> where they are in a polarized, or resting state.
CSV myocytes depolarized	ecgOnto:068	phase	Phase of the <i>CSV myocytes</i> where they are in a depolarized state.
To generate CEI	ecgOnto:069	mode	Function that characterizes the <i>SA node</i> and is realized by the process of <i>depolarization of pacemaker SA node myocytes</i> .
CEI generator	ecgOnto:070	role mixin	General role played by any entity as generating a <i>CEI</i> .
SA node as CEI generator	ecgOnto:071	role	Role played by the <i>SA node</i> as generating a <i>CEI</i> . It specializes the general role of <i>CEI generator</i> .
Depolarization of pacemaker SA node myocytes	ecgOnto:072	complex event	Process that brings the <i>SA node myocytes</i> from a polarized to a depolarized state. This process generates a <i>CEI</i> .
Cardiac electrical impulse	ecgOnto:073	mode	<i>CEI</i> . Electrical wavefront that moves down toward from the <i>CSA</i> to the <i>CSV</i> reaching all the <i>conducting system of the heart</i> , with polarized cells at the front, followed by depolarized cells behind. It inheres in some conductor in order to exist.
To conduct CEI	ecgOnto:074	mode	Function that characterizes both the <i>CSA</i> and <i>CSV</i> , and is realized by either the process of <i>depolarization of CSA myocytes</i> and or by the process of <i>depolarization of CSV myocytes</i> .
CEI conductor	ecgOnto:075	role mixin	General role played by any entity as conducting the <i>CEI</i> .
SA node as CEI conductor	ecgOnto:076	role	Role played by the <i>SA node</i> as conducting the <i>CEI</i> . It specializes the general role of <i>CEI conductor</i> .
CSA as CEI conductor	ecgOnto:077	role	Role played by the <i>CSA</i> as conducting the <i>CEI</i> . It specializes the general role of <i>CEI conductor</i> .
CSV as CEI conductor	ecgOnto:078	role	Role played by the <i>CSV</i> as conducting the <i>CEI</i> . It specializes the general role of <i>CEI conductor</i> .
Depolarization of CSA myocytes	ecgOnto:079	complex event	Process that brings the <i>CSA myocytes</i> from a polarized to a depolarized state. It is a realization of the function <i>to conduct CEI</i> .
Depolarization of CSV myocytes	ecgOnto:080	complex event	Process that brings the <i>CSV myocytes</i> from a polarized to a depolarized state. It is a realization of the function <i>to conduct CEI</i> .
To restore EPs	ecgOnto:081	mode	Function that characterizes the <i>CSV</i> and is realized by the process of <i>repolarization of CSV myocytes</i> .
EPs accumulator	ecgOnto:082	role mixin	General role played by any entity as accumulating electrical potentials.
CSV as EPs accumulator	ecgOnto:083	role	Role played by the <i>CSV</i> as accumulating electrical potentials. It specializes the general role of <i>EPs accumulator</i> .
Repolarization of CSV myocytes	ecgOnto:084	complex event	Process that brings the <i>CSV myocytes</i> from a depolarized to a polarized state. It is a realization of the function <i>to restore EPs</i> .

Table 5: ECG Ontology class dictionary: ECG ontology. Relations and classes of the ECG Ontology appear in the textual definitions emphasized.

Term	Class ID	UFO Type	Textual Definition
Record	ecgOnto:085	kind	ECG data record (in any medium) resulting from a <i>recording session</i> and essentially composed by an ECG <i>waveform</i> .
Recording session	ecgOnto:086	complex event	Medical service in which the <i>patient</i> is subject of ECG recording by some <i>recording device</i> . By integrating two different perspectives, the session can be said to coincide to the <i>observation series</i> .
Recording device	ecgOnto:087	kind	Device used to acquired (to record) an ECG from a given <i>patient</i> by means of <i>electrodes</i> . Also called electrocardiograph.
RD as recorder	ecgOnto:088	role	<i>Recording device</i> as it plays the role of an ECG recorder.
Person	ecgOnto:089	kind	Individual human being.
Patient	ecgOnto:090	role	<i>Person</i> as he/she plays the role of being subject of care, i.e., scheduled to receive, receiving, or having received a healthcare service (based on ISO/TC 18308:2003).
Waveform	ecgOnto:091	subkind	<i>Geometric form</i> (which is-a non-elementary form) constituted by the whole <i>sample sequence</i> resulting from the <i>observation series</i> carried out in the context of an ECG <i>recording session</i> .
Observation	ecgOnto:092	atomic event	Measurement of the p.d. between two regions of the patient's body carried out by an ECG <i>recording device</i> by means of two <i>electrode placements</i> on those regions. The <i>placements</i> are defined according to an ECG <i>lead</i> .
Observation series	ecgOnto:093	complex event	Series of evenly spaced in time <i>observations</i> carried out in an ECG <i>recording session</i> .
Sample	ecgOnto:094	kind	Voltage value resulting from an <i>observation</i> .
Sample sequence	ecgOnto:095	collective	Ordered sequence of <i>samples</i> resulting from ( <i>sample sequence of</i> ) an <i>observation series</i> .
Lead	ecgOnto:096	kind	Viewpoint of the heart activity that emerges from an observation series of the p.d. between two electrode placements on specific regions of the patient's body surface.
Electrode	ecgOnto:097	kind	Electrical conductor <i>part</i> of the <i>recording device</i> to be placed on a specific <i>body surface region</i> of the <i>patient</i> .
Body surface region as object of measure	ecgOnto:098	role	<i>Body surface region</i> as it plays the role of being object of voltage measurement.
Electrode as measurer	ecgOnto:099	role	<i>Electrode</i> as it plays the role of a voltage measurer.
Placement	ecgOnto:100	relator	Physical contact between an <i>electrode</i> and a specific <i>body surface region</i> to measure a voltage value.
Geometric form	ecgOnto:101	category	Form that emerges from the connection of a set of n-tuple values as they are projected into some n-dimensional conceptual (or geometric) space.
ECG form	ecgOnto:102	kind	<i>Geometric form</i> constituted by a given <i>sample sequence</i> .
Non-elementary form	ecgOnto:103	subkind	Any arbitrary <i>ECG form</i> which is not an <i>elementary form</i> .
Elementary form	ecgOnto:104	subkind	<i>ECG form</i> that directly <i>maps</i> a cohesive electrophysiological event in the heart behavior or connects two <i>ECG forms</i> which do it.

Continued on next page

Table 5 – continued from previous page

Term	Class ID	UFO Type	Textual Definition
Cycle	ecgOnto:105	subkind	<i>Elementary form</i> periodically repeated in the ECG waveform that indirectly indicates a heart beat. It is composed by the mereological sum of the <i>P wave</i> , <i>PR segment</i> , <i>QRS complex</i> , <i>ST segment</i> , <i>T wave</i> and <i>baseline</i> . The <i>QRS complex</i> is an <i>essential part</i> of it, as the peak of the <i>R wave</i> is considered a reference point to define it.
Wave	ecgOnto:106	subkind	<i>Elementary form</i> that bears necessarily the property of having a peak.
Segment	ecgOnto:107	subkind	<i>Elementary form</i> that connects two waves and does not have a peak.
P wave	ecgOnto:108	subkind	Wave that maps the electrophysiological process of <i>depolarization of the conducting system of atria</i> .
Q wave	ecgOnto:109	subkind	Wave part of the <i>QRS complex</i> which is connected to both the <i>PR segment</i> and the <i>R wave</i> .
R wave	ecgOnto:110	subkind	Wave that is an <i>essential part</i> of the <i>QRS complex</i> . Its peak is considered a reference point to define a cycle.
S wave	ecgOnto:111	subkind	Wave part of the <i>QRS complex</i> which is connected to both the <i>R wave</i> and the <i>ST segment</i> .
T wave	ecgOnto:112	subkind	Wave that maps the electrophysiological process of <i>repolarization of the conducting system of ventricles</i> .
QRS complex	ecgOnto:113	subkind	<i>Elementary form</i> that maps the electrophysiological process of <i>depolarization of the conducting system of ventricles</i> . It is composed by the mereological sum of the Q, R and S waves, though it has as an <i>essential part</i> only the <i>R wave</i> .
PR segment	ecgOnto:114	subkind	Segment that connects the <i>P wave</i> offset to the <i>QRS complex</i> onset.
ST segment	ecgOnto:115	subkind	Segment that connects the <i>QRS complex</i> offset to the <i>T wave</i> onset.
TP segment	ecgOnto:116	subkind	Segment that connects the <i>T wave</i> offset of a given <i>Cycle</i> to the <i>P wave</i> onset of the next one.
Baseline	ecgOnto:117	subkind	Isoelectric line composed by parts of the <i>Waveform</i> where the heart is not performing any electrophysiological activity.
Normal	ecgOnto:118	subkind	ECG <i>elementary form</i> annotated by a physician or a computer program as matching its expected geometric pattern.
Abnormal	ecgOnto:119	subkind	ECG <i>elementary form</i> annotated by a physician or a computer program as not matching its expected geometric pattern.

## 5.8 Conclusions

In this chapter we have developed an ECG ontological theory which outlines a documented ECG Ontology. The key points worth to remind are:

- The ontological theory proposed here constitutes a solution-independent theory of the ECG. It has been developed in an effort to accurately represent the ECG domain upon a sound ontological basis. For this reason it is grounded in the top-level ontology UFO.

- The ECG Ontology handles *what the ECG is* on both sides of the patient and of the physician. Objectively, the domain it is supposed to represent is defined by means of competence questions. Moreover, the ECG Ont. is fully documented by means of class dictionary.
- The ECG Ontology is coupled to two extra original sub-ontologies, viz., the anatomy for ECG and heart electrophysiology sub-ontologies. It imports the OBO Relation Ontology (RO) for building upon basic relations standardized in the biomedical domain.

The ECG ontological theory developed here has been preliminary reported by us in (130). The “off-line” applicability of the ECG ontology outlined here as a resulting artifact of our a domain analysis is demonstrated in Chapter 7. In what follows, the ECG Ontology is used in a process of design to derive a computable artifact useful for, say, knowledge-based applications. One such application that is fueled by the ECG Ont. implementation is in fact presented in Chapter 8.

## 6 ECG Ontology Implementation

This chapter reports the implementation of the ECG Ontology in an ontology codification formalism. As said before, in our project we have chosen OWL DL and its SWRL extension for that, cf. our rationale in Section 4.2.

This chapter is organized as follows. In Section 6.1 we describe basic design patterns for implementing the ECG Ontology's entities by means of OWL primitives. In the sequel we provide a picture of the ECG OWL Ontology outlined in Section 6.2. This section is not intended to present every OWL class or property implemented. The implementation can however be downloaded at the project website<sup>1</sup>. It is object of evaluation by making use of reasoning services in Section 6.3, and then discussed in Section 6.4. Section 6.5 then provides our final remarks.

### 6.1 Basic Design Patterns

In the design task of transforming ECG Ontology's entities to OWL elements, some intuitive basic design patterns has been used as follows. Recall that the ECG Ontology's classes and relations correspond, respectively, to FOL unary and binary predicates. Besides, keep in mind that in this thesis we write OWL to refer to OWL 1.0 and within the specific OWL family named OWL DL.

- *Classes as OWL Classes*: perhaps the most intuitive design pattern is that of implementing every ECG Ontology class (also represented as FOL unary predicates in our ECG theory) as an OWL class. Notwithstanding, their additional (and noteworthy) distinguishing characteristic of instantiating a specific UFO type is unfortunately not possible to be set in OWL. This is because the UFO types are second-order and as such they would jeopardize our desideratum for efficient automated reasoning in a formalism like OWL. The corresponding UFO types are then set in the OWL classes by means of OWL annotations, which are useful either for human reading or even for computer programs if aware of the annotation syntax. Example: *Record* is implemented as an OWL class with the annotation *rdfs:comment* “*ufoType: kind*”@en.
- *Relations as OWL Object Properties*: the relations used in the ECG Ontology (which are listed in Table 2) can be specified in OWL by means of the so-called *object properties*. It may be worth mentioning that the OWL object properties mentioned here henceforth refer to both the instance- and class-level versions of the ECG Ontology's relations. That is, both of them are collapsed into one OWL object property. Some important meta-properties of binary relations can still be expressed in OWL, viz., symmetry, transitivity, functional and inverse functional. However, all the FOL axiomatization that restricts their interpretations and usage has no room for. The low expressiveness allowable for defining binary relations is one of the most limitations of OWL DL / SWRL, but this price is paid for keeping in favor of holding efficient automated

---

<sup>1</sup><http://nemo.inf.ufes.br/biomedicine/ecg.html>.

reasoning. Besides, an OWL object property allows the developer to set for a given OWL object property its class domain and range. This, however, happens to be very useful only for relations that have a quite specific domain and range in their own right - e.g., *sample sequence of*, defined to hold only between a Sample sequence and an Observation series; whereas general relations like *part of* should not have these properties specified.

- *Datatypes as OWL Datatype Properties*: classes that are *datatypes* (i.e., instances of the UFO type *quality* stereotyped in OntoUML by *datatype*) can be represented as OWL datatype properties. Examples of datatypes are: *age*, projected ontologically into the conceptual structure of natural numbers and then implemented as an OWL datatype property with non-negative integers as range; and *date time*, corresponding to timestamp values like “2038-01-09 03:14:07” which are implemented as an OWL datatype property as well with date/time as range.
- *Asserted Datatype Properties as OWL Datatypes Restrictions*: classes can hold properties that consist of projections into conceptual spaces, named *datatypes* in OntoUML. For instance, the *start time* of a Recording session which is projected into the *Date time* datatype. Such a projection can be represented in OWL by making use of an OWL *datatype restriction*, e.g., *startTime some dateTime* asserted as a restriction for the class Recording session.
- *Asserted Relations as OWL Object Restrictions*: the assertion of relations between classes (e.g., Heart *part\_of* Cardiovascular system) fits in OWL as object restrictions. They are to be set in OWL as a restriction for a given class according to the semantics of the relation in hand. In the example just mentioned, the object restriction *part\_of some Cardiovascular system* should be asserted in the class Heart - the use of *some* has been chosen according to the definition of the class-level *part of*.

An OWL restriction for a given OWL class can be either (i) a necessary condition for this class membership, i.e., any individual supposed to instantiate the class must respect it, or (ii) a necessary and sufficient condition which actually constitute the definition of the class, i.e., the definition of the class membership. FOL formulae can fall in OWL either as such conditions or as SWRL rules. Thus, if a FOL formula can be expressed as a horn-like rule (just as R1 - R20 can), then it can be expressed as a SWRL rule (often with some loss in expressivity). Although a SWRL rule is free of the context of any OWL class, it can serve for asserting a sufficient condition for an OWL class membership. Finally, a SWRL rule can be specified by using unary predicates referring to OWL classes and binary predicates standing for OWL (object or datatype) properties. Variables are used to range over individuals. This can be illustrated by the example that follows. This particular rule denotes a sufficient condition for an individual  $x$  instantiate *Class\_2*.

$$Class\_0(?x) \wedge Class\_1(?y) \wedge associated\_to(?x,?y) \rightarrow Class\_2(?x)$$

## 6.2 The ECG OWL Ontology

According to those design patterns, the ECG Ontology has been implemented in OWL DL / SWRL. We illustrate as follows pieces of the ECG OWL Ontology outlined as a result<sup>2</sup>. For this we make use of the Manchester OWL

<sup>2</sup>The annotations are omitted for the sake of brevity.

syntax (131). It is derived from the OWL Abstract Syntax, but is less verbose and minimizes the use of brackets (132).

As the class implementations are direct, we focus more on the relations and exemplify their use as class restrictions in each of the sub-ontologies. The subclass (or subsumption, is-a) relation is OWL built-in; every subclass assertion depicted in the models presented in Chapter 5 is then directly asserted in OWL as well, see Figure 41.

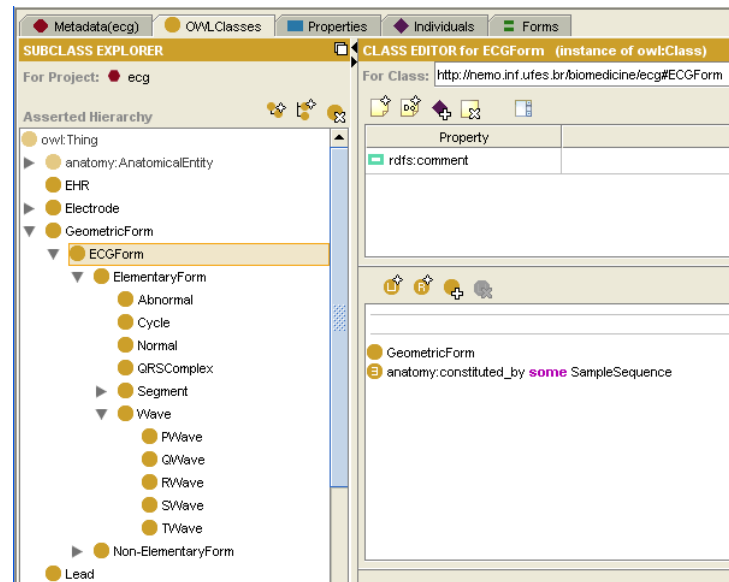


Figure 41: General picture of the ECG OWL Ontology edited in Protege. The nodes on the left are the ECG Ontology’s classes represented as OWL classes. They are organized in a subsumption hierarchy. On the right, two OWL object restrictions are asserted for the class ECG form, viz., *subclass\_of* Geometric form, and *constituted\_by* some Sample sequence. They denote necessary conditions for the ECG form class membership.

### 6.2.1 The OBO RO Extension

The *proper part of* relation and its inverse *has proper part* as we defined in Section 5.2 for the anatomy sub-ontology are already implemented in the OWL codification of the OBO RO available at the project website<sup>3</sup>. They have the labels *proper\_part\_of* and *has\_proper\_part*, respectively. These relations, so as the other RO relations and the extension we have developed here are implemented as keeping not much of their expressivity. In particular, the inseparable and essential distinctions of proper parthood defined in Section 5.2 find no place in this implementation framework because their distinctions require to take account of time (we comment this limitation further on in this text). As discussed above, only some meta-properties can be set for relations in OWL. This impossibility of keeping in an OWL object property the FOL axiomatization its binary relation counterpart is characterized by cannot be ignored. Besides, in our ECG theory there are also relations that come to be ternary in virtue of time arguments. These have been implemented here without considering the time argument. A full discussion on the limitations of our implementation is provided in Section 6.4. As follows, Table 6 provides a summary of the OWL object properties contemplated in the ECG Ontology’s implementation.

<sup>3</sup><http://www.obofoundry.org/ro/>.

Table 6: OWL object properties derived from ECG Ontology’s relations and their features. Except by the two latter (conducted by and mediates), all of them have their inverse counterparts also implemented by holding the same meta-properties.

<i>objectProperty</i>	<i>subPropertyOf</i>	<i>InverseOf</i>	<i>Characteristics</i>
subcollection_of	ro:proper_part_of	has_subcollection	transitive
grain_of	ro:part_of	has_grain	-
constituted_by	-	constitutes	transitive
partially_constituted_by	constituted_by	partially_constitutes	transitive
realized_by	-	is_realization	-
actually_realized_by	realized_by	is_actual_realization	-
characterized_by	ro:relationship	characterizes	-
produced_by	ro:relationship	produces	-
generated_by	ro:relationship	generates	-
conducted_by	ro:relationship	-	-
mediates	ro:relationship	-	-

### 6.2.2 Anatomy OWL Sub-Ontology

The anatomy sub-ontology comprises 62 classes, whose most characteristic restrictions are subsumption and parthood relationships. We illustrate this sub-ontology by putting below an excerpt of it containing three classes, viz., Heart, Conducting system of atria and SA node myocytes.

#### Heart

```
Class: anatomy:Heart
```

```
SubClassOf:
```

```
    anatomy:Organ,
    ro:has_proper_part some anatomy:LeftAtrium,
    ro:has_proper_part some anatomy:LeftVentricle,
    ro:has_proper_part some anatomy:RightAtrium,
    ro:has_proper_part some anatomy:RightVentricle,
    ro:has_proper_part some anatomy:WallOfHeart,
    ro:proper_part_of some anatomy:CardiovascularSystem
```

#### Conducting system of atria

```
Class: anatomy:ConductingSystemOfAtria
```

```
SubClassOf:
```

```
    anatomy:ConductingSystemOfSubdivisionOfHeart,
    roExtension:partially_constituted_by some anatomy:CSAMyocytes,
    roExtension:partially_constituted_by some anatomy:ECMOfCSA,
    ro:has_proper_part some anatomy:ConductingSystemOfRightAtrium,
    ro:has_proper_part some anatomy:LeftBranchOfBachmannsBundle,
```



```
ro:proper_part_of some anatomy:ConductingSystemOfHeart
```

### SA node myocytes

```
Class: anatomy:SANodeMyocytes

SubClassOf:
  anatomy:CellCluster,
  roExtension:has_subcollection some anatomy:PacemakerSANodeMyocytes,
  roExtension:has_subcollection some anatomy:TransitionalSANodeMyocytes,
  roExtension:partially_constitutes some anatomy:SANode,
  roExtension:subcollection_of some anatomy:CSAMyocytes
```

## 6.2.3 Physiology OWL Sub-Ontology

The sub-ontology of physiology is quite shorter, comprising 22 classes. We illustrate this sub-ontology by putting below an excerpt of it containing one example of each of the following sort of entities: myocytes' phases, functions, processes and roles.

### Myocytes' phases

```
Class: CSVMyocytesPolarized

SubClassOf:
  anatomy:CSVMyocytes

DisjointWith:
  CSVMyocytesDepolarized
```

### Functions

```
Class: ToConductCEI

SubClassOf:
  owl:Thing,
  roExtension:characterizes some anatomy:ConductingSystemOfAtria,
  roExtension:characterizes some anatomy:ConductingSystemOfVentricles,
  roExtension:characterizes some anatomy:SANode,
  roExtension:realized_by some DepolarizationOfCSAMyocytes,
  roExtension:realized_by some DepolarizationOfCSVMyocytes
```

### Processes

```
Class: DepolarizationOfCSAMyocytes

SubClassOf:
  owl:Thing,
  roExtension:is_realization some ToConductCEI,
  ro:has_participant some CSAAsCEIConductor,
  ro:has_participant some CardiacElectricalImpulse
```

### Roles

```

Class: SANodeAsCEIGenerator

SubClassOf:
    anatomy:SANode,
    CEIGenerator,
    ro:participates_in some DepolarizationOfPacemakerSANodeMyocytes

```

## 6.2.4 ECG OWL Sub-Ontology

Finally, the ECG sub-ontology imports the two previously presented and contains 34 classes. Two OWL object properties are implemented in this sub-ontology, namely *sample sequence of* and *maps*.

### Relation *sample sequence of*

```

ObjectProperty: sample_sequence_of

Characteristics:
    Functional

Domain:
    SampleSequence

Range:
    ObservationSeries

```

### Relation *maps*

```

ObjectProperty: maps

Characteristics:
    Functional

Domain:
    ECGForm

```

If a given relation holds the property of being *functional* then each instance which stand in its domain bear this relation with at most one instance in its range. For example, a Person has at most one value for the datatype property “age”, while an Employee can have in a given conceptualization at most one Direct supervisor. Accordingly, a Sample sequence individual is *sample sequence of* some and at most one Observation series individual. The *maps* relation in turn, although being functional as well and restricted in its domain, has not been restricted in its range. The rationale for that is to let open the world for further extensions with respect to electrophysiological processes possibly mapped by other ECG forms.

In addition, in the ECG sub-ontology a number of datatype properties take place. We illustrate this by means of the properties *start* and *p.d.* as follows.

### Start

```

DataProperty: start

Characteristics:
    Functional

Domain:
    RecordingSession

```

```

Range:
    dateTime

```

### p.d.

```

DataProperty:  p.d.

Characteristics:
    Functional

Domain:
    Sample

Range:
    float

```

Furthermore, among the 34 classes present in the ECG sub-ontology, we draw attention to the implementation of three of them, viz., Record, ECG form and Sample sequence.

### Record

```

Class:  Record

SubClassOf:
    owl:Thing,
    roExtension:produced_by some RecordingSession,
    ro:has_proper_part some Waveform

```

### ECG form

```

Class:  ECGForm

SubClassOf:
    GeometricForm,
    roExtension:constituted_by some SampleSequence

```

### Sample sequence

```

Class:  SampleSequence

SubClassOf:
    owl:Thing,
    sample_sequence_of some ObservationSeries,
    roExtension:constitutes some ECGForm,
    roExtension:has_grain some Sample,
    p.d.sequence some string,
    sample_rate some int

```

## 6.2.5 FOL Formulae as OWL Restrictions and SWRL Rules

By considering the FOL formulae F1 - F20 presented in Section 5.6, we have been able to codify in OWL DL / SWRL those that are not strictly dependent on time arguments, except by F11 which cannot be expressed as a

Horn-clause. An application-independent implementation of F3, F5b<sup>4</sup>, F8, F12, F13, F15, F18 and F20 would require to take in account the time arguments used in assertions like “*c* located\_in *c*<sub>1</sub> at *t*”, which is not the case in the present implementation.

We adopt in this implementation a perspective in which all temporally extended entities such as Depolarization of CSV work *as if* have been occurred in its entirety already. That is, if a given process is instantiated by some individual, then their participants have been participated already in the process. This in fact makes sense in terms of real world. The ECG data can only be considered after all processes’ manifestation, since (i) the electrophysiological ones occur in between the ECG recording session, i.e., while the ECG does not exists yet; and (ii) the recording session process itself is that which actually produces the ECG, which therefore starts to exist at the last time instant of this process. However, we cannot either assert or infer that a given individual “*c* is located\_in *c*<sub>1</sub> at *t*” and in *c*<sub>2</sub> at *t*’” if not by working with a temporal knowledge base. In our design phase we have refrained from this mostly in virtue of our project’s time constraints. We could not get further into an investigation towards a proper implementation framework (cf. discussion in Section 9.3) for tackling these time arguments. A direct consequence of this, as commented in the next section, is that some CQs could not be answered by automated reasoning. This matter is still discussed in Chapter 9.

Now we refer to the formulas implemented genuinely, either as OWL class restrictions (viz., F1, F2, F6, F7, F14, F16, F17) or SWRL rules (viz., F4, F5a, F9, F10, F19). Starting by the former set, the formulas F1, F6 and F16 refer to mapping relations which are necessary conditions for membership of waves’ classes. They are part of the ECG sub-ontology and are implemented in OWL as follows.

### Mapping relations: F1, F6 and F16

```

Class: PWave

SubClassOf:
    Wave,
    maps some physiology:DepolarizationOfCSAMyocytes,
    ro:proper_part_of some Cycle,
    offset some int,
    onset some int,
    peak some int

DisjointWith:
    QWave,
    SWave,
    RWave,
    TWave

Class: QRSComplex

SubClassOf:
    ElementaryForm,
    maps some physiology:DepolarizationOfCSVMMyocytes,
    ro:has_proper_part some RWave,
    ro:proper_part_of some Cycle,
    offset some int,
    onset some int,
    peak some int

```

<sup>4</sup>The formula F5 has been partitioned into two formulas, one concluding that the CEI has been conducted by the atria (F5a) and another entailing it is located in the ventricular part of AV node at a given time instant (F5b).

```

    DisjointWith:
        Segment

Class: TWave

    SubClassOf:
        Wave,
        maps some physiology:RepolarizationOfCSVMycytes,
        ro:proper_part_of some Cycle,
        offset some int,
        onset some int,
        peak some int

    DisjointWith:
        QWave,
        SWave,
        RWave,
        PWave

```

The remaining formulas F2, F7, F14 and F17 comprise in turn participation relations which are necessary conditions for membership of electrophysiological processes' classes. They are part of the sub-ontology of physiology and are implemented in OWL as follows.

#### Relations of participation: F2, F7, F14 and F17

```

Class: DepolarizationOfCSAMycytes

    SubClassOf:
        owl:Thing,
        roExtension:is_realization some ToConductCEI,
        ro:has_participant some CSAAsCEIConductor,
        ro:has_participant some SANodeAsCEIConductor,
        ro:has_participant some CardiacElectricalImpulse

Class: DepolarizationOfCSVMycytes

    SubClassOf:
        owl:Thing,
        roExtension:is_realization some ToConductCEI,
        ro:has_participant some CSVAsCEIConductor,
        ro:has_participant some CardiacElectricalImpulse

Class: DepolarizationOfPacemakerSANodeMycytes

    SubClassOf:
        owl:Thing,
        roExtension:is_realization some ToGenerateCEI,
        ro:has_participant some CardiacElectricalImpulse,
        ro:has_participant some SANodeAsCEIGenerator

Class: RepolarizationOfCSVMycytes

    SubClassOf:
        owl:Thing,
        roExtension:is_realization some ToRestoreEPs,
        ro:has_participant some CSVAsEPsAccumulator

```

Let us now consider the second set of formulae F4, F9, F10 and F19, which happen to fit as SWRL rules as follows. Notice that F10 is combined to F7 in order to restrict the continuants considered in the antecedent of the formula.

#### SWRL Rules: F4, F5a, F9, F10 and F19

```

F4. physiology:DepolarizationOfCSAMyocytes(?p)  ∧  physiology:ToConductCEI(?f)
    ∧  roExt:realized_by(?f, ?p)  ∧  ecg:Pwave(?c)  ∧  ecg:Normal(?c)  ∧  ecg:maps(?c, ?p)
    →  roExt:actually_realized_by(?f, ?p)

F5a. physiology:DepolarizationOfCSAMyocytes(?p) ∧ physiology:ToConductCEI(?f)
    ∧  roExt:actually_realized_by(?f, ?p) ∧ CEI(?c1) ∧ physiology:CSAAsCEIConductor(?c2)
    ∧  oboRo:has_participant(?p, ?c1) ∧  oboRo:has_participant(?p, ?c2)
    →  roExt:conducted_by(?c1, ?c2)

F9. physiology:DepolarizationOfCSVMMyocytes(?p)  ∧  physiology:ToConductCEI(?f)
    ∧  roExt:realized_by(?f, ?p)  ∧  ecg:QRSComplex(?c)  ∧  ecg:Normal(?c)  ∧  ecg:maps(?c, ?p)
    →  roExt:actually_realized_by(?f, ?p)

F10. physiology:DepolarizationOfCSVMMyocytes(?p)  ∧  physiology:ToConductCEI(?f)
    ∧  roExt:actually_realized_by(?f, ?p) ∧  physiology:CardiacElectricalImpulse(?c1)
    ∧  physiology:CSVAsCEIConductor(?c2)  ∧  oboRo:has_participant(?p, ?c1)
    ∧  oboRo:has_participant(?p, ?c2)
    →  roExt:conducted_by(?c1, ?c2)

F19. physiology:RepolarizationOfCSVMMyocytes(?p)  ∧  physiology:ToRestoreEPs(?f)
    ∧  roExt:realized_by(?f, ?p)  ∧  ecg:TWave(?c)  ∧  ecg:Normal(?c)  ∧  ecg:maps(?c, ?p)
    →  roExt:actually_realized_by(?f, ?p)

```

## 6.3 ECG Ontology Evaluation

The ECG Ontology implementation has been object of an evaluation in terms of consistency, efficiency and competence. For that we have made use of the automatic reasoner Pellet (133), by means of its Java API version 2.0 RC5 in combination to the OWL API Jena. Pellet is used here for checking the logical consistency of the ontology, as well as retrieving information about individuals and their relationships. We then have conducted tests, such as (i) executing the inference procedure and (ii) listing individuals<sup>5</sup>. This is because while (i) can validate the logical consistency of the ontology<sup>6</sup>, (ii) verify the efficiency it is able to afford for information retrieval. Table 7 provides an account of such evaluation carried out for the ECG Ontology implementation.

As a result of these tests, we can conclude that the ECG OWL Ontology: (i) is logically consistent; (ii) affords efficient automated reasoning and information retrieval.

### DL Family

Without considering the SWRL rules, the DL family which the ECG Ontology implementation corresponds to is SHIF(D)<sup>7</sup>. This stands for: (i) S - an abbreviation for ALL; just AL means Attributive Logic: Conjunction, Universal Value Restriction, Limited Existential Quantification, but extended with another L means the inclusion

<sup>5</sup>We have filled in the ontology with individuals as reported in Section 8.1.

<sup>6</sup>If the reasoner successfully return from the inference procedure then it has completed the proof for the model validation.

<sup>7</sup>This DL family is even more computationally tractable than the one that mirrors OWL DL, viz., SHOIN(D). Since SHIF(D) is a subset of SHOIN(D), our initial intend of keeping the implementation expressiveness under the rubric of OWL DL for the the sake of “ensuring effective automated reasoning” (cf. Section 4.2) is attained.

Table 7: Evaluation results for the ECG Ontology implementation. The tests have been carried out on a Microsoft Windows machine featuring AMD Turion 1.8 GHz and 1GB of main memory, in Java 1.6.0\_02, by using Jena-2.5.7 and Pellet-2.0-RC5. The timing measurements have been obtained from performing each of the two tests for 10 times. The procedure that lists all ontology individuals has been called always after the ontology model's validation.

<i>ECG Ontology Metrics</i>	
Class count	129
Object property count	51
Data property count	16
Individual count	311

<i>Timing Measurements in (ms)</i>			
<b>Procedure</b>	<b>Mean</b>	<b>Median</b>	<b>Standard deviation</b>
Make inference	430	383	182
List all individuals	150	78	223

of Complement (for covering also Disjunction and Full Existential Quantification); (ii) H - Role hierarchy (subproperties); (iii) I - Inverse properties; and (iv) (D) - Datatypes; and F - Functional properties.

#### *Competence Questions*

Finally, we consider the ECG Ontology's CQs in this evaluation. Let us take CQ1 to exemplify the strategy we have adopted for verifying how the CQs are addressed. CQ1 states that every **Record** individual must has a **Waveform** individual as a proper part<sup>8</sup>. It is implemented as a simple OWL object restriction for the class **Record**. As such, it consists of a necessary condition for this class membership which is therefore warranted for each **Record** individual. In other words, if any individual of this class does not have a **Waveform** individual as part, the reasoner cannot complete the proof for the model validation and then does not return in efficient time. Ergo, we reflect on this example to draw attention to the fact that *CQs implemented as necessary conditions are proved to be answered if the required individuals exist in the facts base and the reasoner completes the ontology validation*.

By applying this rule, we have then verified the effectiveness of our CQs. The verification of CQ3 has been warranted likewise. For the rest of them except by CQ2, notice that their FOL axiomatization (cf. Section 5.7.1) is covered by the formulae described in Section 5.6, which are implemented as reported above. As said before, however, some of them (viz., CQ6 and CQ7) require to cope with time arguments which have not been able to be considered in our implementation. CQ2 lies in time-dependent predicates likewise. We then recognize that as an issue that restricts to some extent the automated evaluation of the ECG Ontology's completeness. Nevertheless, in spite of this limitation of our implementation framework, much of the ECG Ontology competence has been empirically verified.

## 6.4 Discussion

As we have seen, a significant part of the ECG Ontology's axiomatization could not be implemented in OWL DL / SWRL. This is in virtue of the tradeoff between expressiveness and computational tractability well-known in Knowledge Representation (55). We, however, along the lines traced in Section 2.3, refer to the artifact reported

<sup>8</sup>Actually, as an essential proper part; however, as said before this stronger parthood could not be preserved in the ontology implementation.

in this chapter not as an ECG ontology, but rather as a partial version of it strictly designed to be computationally tractable. This artifact can be called the *lightweight ECG ontology*, in contrast to the *reference ECG ontology*. In face of this, a question that is often raised in ontology communities is the following: if not much of the reference ontology axiomatization can be preserved in the ontology codification, why not to start directly from the implementation? Or, to put it differently by referring to a usual expression, “why do it the hard way”? We then refer once more to Thomas Bittner’s presentation in Rome 2005<sup>9</sup>, to echo that,

*“because this is the only way to produce good ontologies.”*

In this sense, we sustain that a lightweight ontology can be hardly good if not derived from a reference one. We list below some points we have developed in the course of this thesis that provide support for this claim.

- The backbone subsumption taxonomy of a given ontology benefits a lot from being developed by following a methodology principled on ontological foundations such as OntoClean. In this way, even though most of the classes’ meta-properties cannot be set in the lightweight ontology implementation, they would have been used to define the taxonomy itself.
- So do mereological relations that are further distinguished among their several kinds, viz. (34, Chapter 5): grain of / member of, subcollection of, essential / inseparable / shareable part of and so on. Even though these mereological distinctions often have no room left to in the ontology implementation, they can contribute to keep ontological soundness if captured as conceiving the reference artifact. For one, consider how an understanding of the distinction between *grain of* and general *part of* has been meaningful to avoid illegitimate transitivity propagation through different levels of granularity in the ECG Ontology.
- Albeit almost nothing of the axiomatization of binary relations can be represented in a lightweight ontology, understanding them in depth is still purposeful in order to, (i) first, assert them as consciously as possible between the domain universals by building upon how they manifest between their instances; (ii) second, establish necessary, sufficient, or both conditions for class membership as consequence of the (well-axiomatized and understood) relations a given universal stand for.

As we have seen, all this has been applied in the development of the ontology proposed in this thesis. We then feel comfortable enough to state that even the (much less expressive) ECG OWL Ontology is well-founded - in the sense that it is founded on formal ontological principles. Our point here is not other than that the expressiveness/tractability tradeoff does not take away the benefits of lightweight ontologies which happen to be well-founded.

## 6.5 Conclusions

In this chapter we report an implementation of the ECG Ontology in the ontology codification language OWL DL / SWRL. This favors one of our objectives which is seeking ontology integration w.r.t. both the OBO foundry and the semantic web effort. The ECG OWL ontology can be downloaded at our project website<sup>10</sup>. Overall, the contents of this chapter can be summarized as follows.

<sup>9</sup><http://ontology.buffalo.edu/05/wg6/bittner.ppt>. Accessed on April 05, 2009.

<sup>10</sup><http://nemo.inf.ufes.br/biomedicine/ecg.html>.



- The ECG Ontology implementation reported here has been object of an evaluation. Our tests have verified the logical consistency of the ontology as well as its effectiveness for information retrieval. Furthermore, this implementation has been the source for verifying that the ontology's CQs (without considering time-dependent predicates) are covered. In other words, it has been warranted the ECG Ontology competence.
- In virtue of the low expressiveness of OWL DL / SWRL in comparison to FOL, part of the ECG Ontology axiomatization could not be expressed in its implemented version.
- Nevertheless, we contend that striving for a (strongly-axiomatized) reference ontology is unavoidable in order to produce a good ontology. Benefits reached with an ontology engineering process principled in Formal Ontology are preserved even in the implemented lightweight ontology.

The OWL DL / SWRL adapted version of the ECG Ontology comes to be susceptible for effective automated reasoning. This is also demonstrated in Chapter 8 in an “on-line” reasoning-based application of the ECG Ontology.

## 7 Application in Conceptual Modeling

This chapter presents an application of the ECG Ontology in the field of Conceptual Modeling. We apply here the ECG Ontology to foster interoperability in Health Informatics among ECG data format standards which are currently *in use*. The contents of this chapter result also from our experience in dealing with ECG data standards in virtue of the TeleCardio project (134), for that we developed a genuine ECG data format (135).

We start in Section 7.1 with an introduction to the most referred ECG data standards from a historical perspective. We then present each of these standards in what concerns their basic data format characteristics in Section 7.2. Subsequently, Section 7.3 discusses how an ontology can be used to foster interoperability in Conceptual Modeling. Finally, in Section 7.4 we present an integration experiment between the ECG Ontology and the ECG standards contemplated here. We then report our conclusions in Section 7.5.

### 7.1 ECG Data Standardization: An Ongoing Story

Electrocardiography became clinically feasible with Einthoven's invention of the string galvanometer - for which he received the Nobel Prize, in 1902. Since then the measurement of bioelectric potentials generated in the human heart has been object of research by biomedical engineers leading then to many improvements in instrumentation and digital computers. In recent years, the latest information and communication technologies have set the ground for the emergence of Telecardiology (136), which relies primarily on the transmission of the ECG. The storage and transmission of ECG records have then been object of several initiatives regarding standardization. They aim at a suitable digital ECG data format mostly for: (i) supporting continuity of care by maintaining such ECG records in electronic health records (EHR)<sup>1</sup>, and (ii) easily communicating the cardiac test results between the various health care providers.

The AHA/MIT-BIH and SCP-ECG standards were conceived by aiming at the storage and transmission of ECG records, respectively. Thereafter, other ECG standards were created as bearing the feature of being XML-based. One of the motivations was to meet an arising requirement at the moment: the need for flexibility and interoperation in the context of the Internet. Namely, they are FDA XML (or just FDADF) and HL7 aECG. All of these standards can be said reference ones in what concerns ECG data format in Health Informatics. We briefly tell their stories in what follows.

#### *AHA/MIT-BIH*

Since 1975, the Boston's Beth Israel Hospital (BIH, now the Beth Israel Deaconess Medical Center) and MIT

---

<sup>1</sup>Electronic Health Record. "A repository of information regarding the health of a subject of care [(a *patient*)], in computer processable form" (ISO/TC 18308:2003).

(Massachusetts Institute of Technology) have carried out joint research on the analysis of physiological medical exams (137). The first result, deployed in the early 80's, was the MIT-BIH Arrhythmia Database. This is a thoroughly tested and standardized resource for the detection and evaluation of cardiac arrhythmias which has been used in cardiac physiology research around the world (138). Towards the same direction, at that time the American Heart Association (AHA) was deploying the AHA Database for Evaluation of Ventricular Arrhythmia Detectors.

Posteriorly, the Research Resource for Complex Physiologic Signals project was launched by researches from the BIH, MIT, Harvard Medical School, Boston University, and McGill University. This multi-institutional project was sponsored by the National Center for Research Resources (NIH/NCRR) by aiming at providing support for the ongoing research as well as to set new topics concerning complex physiological signals. As a result of that project, three resources were deployed (137):

- *PhysioNet* (137): a website for providing the biomedical scientific community with a free access to physiological data and corresponding software for accessing them.
- *PhysioBank*: an open access database containing over 4000 vital sign records (mostly ECGs), for which many of them are annotated. This database is available in *PhysioNet*.
- *PhysioToolkit*: a software toolkit for the visualization, importing/exporting data, signal analysis and simulation. Also available in *PhysioNet*.

In 1990, the standard principles employed in the *PhysioNet* resources were extended to aggregate the European ST-T Database (ESC DB). The *PhysioNet* have also been leaving room databases of other physiological signals such as blood pressure, breathing, oxygen saturation and electroencephalogram. For the purpose of this work, however, it is worth to mention that the *PhysioNet* data (including the ECG data) is available in flat files that follow the AHA/MIT-BIH standard nomenclature, structure and so on. We touch on this point further on in this text. Besides, we make use in this thesis in Chapter 8 of ECG data available at PhysioNet as input for a knowledge-based application.

### **SCP-ECG**

The Standard Communications Protocol for Computer-Assisted Electrocardiography (SCP-ECG) is devoted to specify a data format and a transmission procedure for ECG records. First of all, from 1989 to 1990, the SCP-ECG started to be designed by seeking a novel ECG compression method to advance former techniques used. There was a cooperation among european, american and japonese manufactures and users. The new method developed was then referred to as warranting quality of service (QoS) (23).

The European Committee for Standardization (CEN) then approved SCP as the pre-standard ENV 1064 in 1993. Subsequently, it became an ISO recommendation identified as ISO TC215, being then constantly updated by the working groups WGI, WGII, WGIII and WGIV. At this point it is already approved as ISO/DIS 11073-91064 (23). This text, however, is based on the SCP-ECG version 2.1 (prEN 1064:2005) published in 2004 by CEN/TC-251 (139). It is still worth mentioning that, in 2002 an european project carried out by industrial parties, health professionals, standardization bodies and so forth was launched for fostering the adoption of SCP-ECG. The openECG project (23) main goals have been, first, (i) to promote a consistent use of data format and transmission standards for ECG records; and second, (ii) to guide the development of similar standards for stress ECG and Holter ECG / real-time monitoring.

### ***FDA XML***

The Center for Drug Evaluation and Research of the Food and Drug Administration (FDA CDER)<sup>2</sup>, as the name suggests, aims at supporting and controlling the safety of drug developers in the USA. The FDA CDER also deals with medical data submission and for this reason has contributed to the development of an ECG XML-based data format. In seeking the ease of vital sign data submission and further analysis, the FDA CDER investigated former ECG standards like the two aforementioned and chose to adopt XML as a data format specification framework. This decision was actually based on the third version of the Health Level Seven (HL7)'s recommendation, as told by N. Stockbridge and B. Brown in (140).

Thus, in April 2002, the FDA XML Data Format (FDADF) was designed for specifying a data format for ECG data and ensure FDA's stakeholders to share it. The FDA XML is proposed with a main contribution by Barry Brown in (24). That document covers not only the ECG data format itself, but also data submission information relevant for message exchange. At this point, however, the FDA XML data format seems to be no distinguished anymore to the aECG standard deployed by the HL7 as described in the following.

### ***HL7 aECG***

The Health Level Seven Inc. (HL7 - <http://www.hl7.com/>) is a non-profit standards developing organization that is accredited by the American National Standards Institute (ANSI). The Annotated ECG (aECG) HL7 standard was created in response to the FDA's digital ECG initiative. It in fact can not be distinguished to the FDA XML itself; since the FDA, sponsors, core laboratories, and device manufactures worked together within HL7 to create an ECG data format standard to meet their needs (141, 140). The aECG standard properly was created by HL7's Regulated Clinical Research Information Management (RCRIM) and accepted by ANSI in May, 2004. The January, 2004 version is then the format the FDA expects to receive all annotated ECGs in (141). This format comprises some of the pieces HL7 had developed to describe other clinical data acquisition settings, but also incorporating added elements necessary to describe ECG waveforms and annotations.

## **7.2 The Reference ECG Data Formats**

The data formats are supposed to be introduced in the same spirit and by the same jargon employed by their maintainers. It is important to seek fidelity as much as possible since we intend to cope with the subtleties that underlie the data models. We then extract (or rather, "excavate") the ECG conceptualization underlying them. Such an "excavation" is based on: their global textual descriptions or definitions (when existing); and the data models properly, which are structured always in a specific data format (e.g., binary format, XML, etc). It is still worth to mention that this task turns out to be not that easy as those ECG standards have no commitment to be comprehensible in a "conceptual-level", but only for implementation. For expressing the structure we assign for each of them, we convey a tree-based schema and a conceptual model in standard UML. The tree-based schema is contemplated here because all of the ECG standards fit the tree-like composition hierarchy. By these means, we can firstly convey a more direct specification of them, to proceed further to bring in a more proper conceptual model in standard UML.

---

<sup>2</sup><http://www.fda.gov/cder/>

### 7.2.1 AHA/MIT-BIH

The AHA/MIT-BIH textual description that follows is an excerpt of the WFDB programming's guide (version 10.4.19) developed by George Moody (142).

The ECG data is available in *PhysioNet* always as part of one of the databases in *PhysioBank*. The databases contain ECG records, each composed of a set of flat files. Each *record* contains a continuous recording from a single subject, and is usually distributed into three files: a header, a data file and an annotation file, which are identified as such by their extension (viz., “.hea”, “.dat” and “.atr”, respectively). For example, the MIT-BIH Arrhythmia Database includes “record 100”, which is composed by the three files “100.atr”, “100.dat”, and “100.he”. The data (or signal) file is in binary format for saving time access and space. It lies in the digitized samples of one or more signals and can be very large. The header file is a short text file that describes the signals, including: the name or URL of the signal file, storage format, number and type of signals, sampling frequency, calibration data, digitizer characteristics, record duration and starting time. At last, the annotation file is often included in the record and is also in binary format. Annotation files contain sets of labels (the annotations), each of which describes a feature of one or more signals at a given time instant in the record. The file “100.atr” cited above, for example, contains an annotation for each QRS complex (that indicates a heart beat) in the recording, indicating its location (time of occurrence) and type (normal, ventricular ectopic, etc.), as well as other annotations that indicate changes in the predominant cardiac rhythm and in the signal quality. Several other annotations are used in other databases in *PhysioBank* to mark other features of the signals.

Signals are commonly understood to be functions of time obtained by observation of physical variables. In AHA/MIT-BIH, a *signal* is defined more restrictively as a finite sequence of integer *samples*, usually obtained by digitizing a continuous observed function of time at a fixed *sampling frequency* expressed in Hz (samples per second). The time interval between any pair of adjacent samples in a given signal is a *sample interval*; all sample intervals for a given signal are equal. The integer value of each sample is usually interpreted as a voltage, and the units are called analog-to-digital converter units, or *adu*. The gain defined for each signal specifies how many adus correspond to one *physical unit* (usually one millivolt, the nominal amplitude of a normal QRS complex on a body-surface ECG lead roughly parallel to the mean cardiac electrical axis). All signals in a given record are usually sampled at the same frequency, but not necessarily at the same gain. For instance, MIT DB records are sampled at 360 Hz; AHA and ESC DB records in turn are sampled at 250 Hz. The sample number is an attribute of a sample, defined as the number of samples of the same signal that precedes it; thus the sample number of the first sample in each signal is zero. Within AHA/MIT-BIH, the units of time are sample intervals; hence the “time” of a sample is synonymous with its sample number.

MIT DB records are each 30 minutes in duration, and are *annotated* throughout; by this we mean that each beat (QRS complex) is described by a label called an annotation. Typically an annotation file for an MIT DB record contains about 2000 beat annotations, and smaller numbers of rhythm and signal quality annotations. AHA DB records are either 35 minutes or 3 hours in duration, and only the last 30 minutes of each record are annotated. ESC DB records are each 2 hours long, and are annotated throughout. The “time” of an annotation is simply the sample number of the sample with which the annotation is associated. Annotations may be associated with a single signal, if desired. Like samples in signals, annotations are kept in time and signal order in annotation files. No more than one annotation in a given annotation file may be associated with any given sample of any given signal. There may be many annotation files associated with the same record, however; they are distinguished by annotator names. The annotator name ‘atr’ is reserved to identify reference annotation files supplied by the developers of the

databases to document correct beat labels.

From this textual description provided above, we take the job of assigning a conceptual structure to the AHA/MIT-BIH. The result of our endeavor is presented in Figure 42 (a data model) and Figure 43 (a conceptual model). Note that this ECG standard do not consider the subject (i.e., the patient) the ECG is acquired from. The reason is that main aspect of interest for this standard is the ECG signal, such that the ECG records in AHA/MIT-BIH are often actually excerpts of original records resulted from recording sessions.

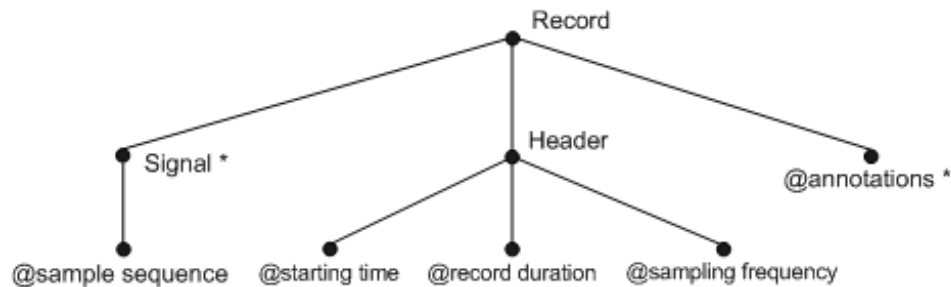


Figure 42: Tree-based data model of the AHA/MIT-BIH. The black circles denote nodes and the '@' symbol indicates leaf elements.

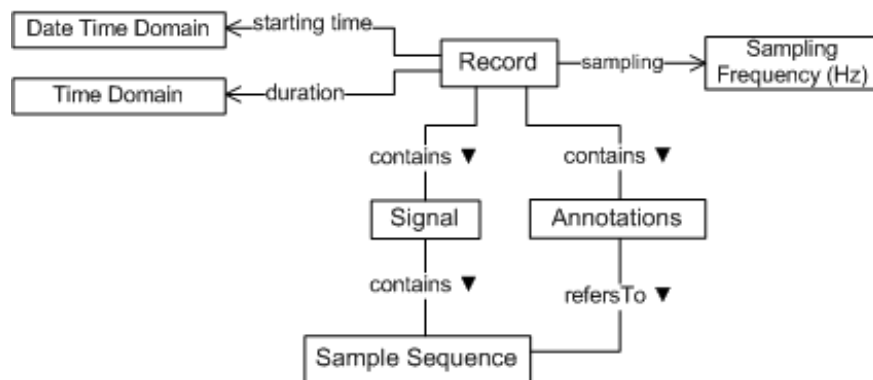


Figure 43: Conceptual model of the AHA/MIT-BIH.

The AHA/MIT-BIH's sample number definition seems to be strongly influenced by the programming language C, which is adopted for the development of the PhysioNet resources. Starting the indexing of array data structures at zero is in fact preferable in some contexts for many programming-motivated reasons<sup>3</sup>; but which fall short in providing a sound ontological basis to think of the first sample made by a device as being the sample "zero<sup>th</sup>". This is one of the several examples that can be found in this ECG standard and in those we are about to introduce that provide evidence of how technological issues guide their conceptualizations.

## 7.2.2 SCP-ECG

The SCP-ECG textual description that follows is an excerpt of the SCP document (version N02-15) developed by the CEN/TC-251 (139).

We have selected what to consider in SCP-ECG. We have left behind aspects of: (i) bit schemata specification, (ii) signal processing and filtering issues, (iii) a number of measurements varying in sort, which even though useful,

<sup>3</sup>This is in fact well-justified by Dijkstra in "Why numbering should start at zero". Confer <<http://www.cs.utexas.edu/users/EWD/transcriptions/EWD08xx/EWD831.html>>. Access on March, 2009.

can be derived from the basic annotations; (iv) issues of the compression method used for encoding/decoding SCP-ECG messages. Besides, we do not consider here aspects that fall actually into the scope of EHR regimes like: medical history, the inclusion of other vital signs (e.g., blood pressure). Diagnostic matters, on the other hand, could be said to be more pertinent as it concerns what the ECG is on the side of the physician. However, getting into this subject is a quagmire that we are not willing to tackle since it calls for pathological aspects of the heart electrophysiology. That is, something that goes beyond the scope of this thesis. In sum, all those issues are important, of course, but not necessary for a task concerned only to the essence of the ECG.

An ECG record structured according to SCP-ECG is divided into sections. The contents of most of them, however, comprise the issues we have left out of discussion here. Table 8 then provides us with the description of sections of our interest.

Table 8: Sections of a SCP-ECG record and their descriptions.

Section	Contents
1	This section contains information of general interest concerning the patient (e.g. patient name, patient ID, age, etc.) and the ECG (acquisition date, time, etc.). This section is mandatory.
3	This section specifies which ECG leads are contained within the record. This section is optional.
4	If reference beats are encoded, then this section shall identify the position of these reference beats relative to the “residual” signal contained in Section 6 below. This section is optional.
5	Reference beats for each lead are encoded if the originating device has identified those complexes. This section is optional.
6	This section contains the “residual” signal that remains for each lead after the reference beats have been subtracted, or if no reference beats have been subtracted, the entire rhythm signal. This section is optional.
8	This section contains the latest actual text of the diagnostic interpretation of the recorded ECG data, including all overreadings if performed. Only the text of the most recent interpretation and overreading shall be included in this section.

Some SCP-ECG term textual definitions are listed below:

- *section*: aggregate of data elements related to one aspect of the electrocardiographic recording, measurement or interpretation.
- *acquiring cardiograph*: cardiograph recording the original ECG signal.
- *record*: entire data file which has to be transmitted, including the ECG data and associated information, such as patient identification, demographic and other clinical data.
- *reference beat*: reference/representative ECG cycle computed through any (but not specified) algorithm comprising the P, QRS and the ST-T waves.
- *rhythm data*: full original ECG data, or the decompressed and reconstructed ECG data at reduced resolution. Rhythm data is typically 10 s in length.

Besides, let us report here pieces of information extracted from the SCP document that provide evidence for the data model we are about to present. In pages 17 and 18, the contents of section 1 are reported in more detail:

“Header information - Patient data / ECG acquisition data - Section 1”. The header information includes: patient identifier, date of acquisition, time of acquisition and acquiring device [or acquiring cardiograph] identification.

In pages 37 and 38 in turn we find: “ECG lead definition - Section 3. This section defines the leads that are transmitted, together with some general administrative information. The detailed information for each lead is as follows. Byte Contents: (i) 1 to 4 (Unsigned) - Starting sample number; (ii) 5 to 8 (Unsigned) - Ending sample number; (iii) 9 - Lead identification.”. From this it is possible to grasp that the “electrode configuration code” is used to set the “lead identification”.

In addition, in page 12, “ECG samples are indexed and numbered starting with sample number 1. Sample index 0 is not used in the present document. The sample index is a ones-based 16-bit index. The first sample starts at time 0.”. In page 40, “The sample numbering shall start with sample number 1 and refers to all leads recorded simultaneously. In order to convert these values to time, the sampling rate of the proper data section [...] should be consulted.”. This sentence finally brings in a basic notion in the ECG recording process, that of *sampling rate*.

All of the description above results from our very task of collecting meaningful information (at a conceptual level) regarding the contents of an ECG record structured in the SCP format. We then present in Figure 44 and Figure 45, respectively, a data and a conceptual model of the SCP-ECG as a result.

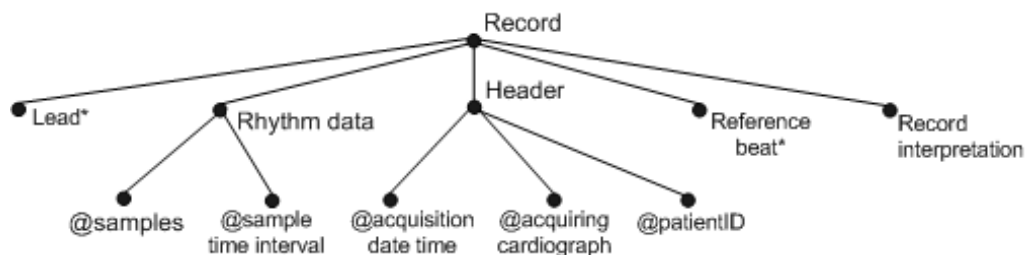


Figure 44: Tree-based data model of the SCP-ECG. The black circles denote nodes and the ‘@’ symbol indicates the leaf elements. An ECG record in the SCP format is composed by several *sections*. The sections containing relevant elements are listed as follows. One or more *leads* are identified in section 3 as being recorded to result in the *rhythm data* of section 6. This data comprises *samples* ordered as they are acquired according to a *sample interval*. Section 1 lies in a *header* which contains the record *acquisition date / time* and an identification of the *acquiring cardiograph* and the *patientID*. Section 4 and 5 are devoted to encode *reference beats* if they are recognized by the acquiring device. Finally, section 8 bears the *record interpretation* made further by a physician and incorporated in the record.

### 7.2.3 FDA XML / HL7 aECG

The FDA XML / HL7 aECG textual description that follows is an excerpt of the HL7 aECG implementation guide developed by Barry Brown and Fabio Badilini and released in March 21, 2005 (141).

Among the standards contemplated here, the FDA XML / HL7 aECG is the one more committed to be comprehensible for human reading. Nevertheless, much of the aECG content is akin to messaging protocols, organizational health policies and other issues related to the ECG recording sessions. These aspects are left out our scope, as they are not strictly related to the exam itself, but actually to broader Healthcare procedures.

ECG data from each lead can be represented using the raw output of its analog-to-digital converter and parameters for scaling and offset. Sets of leads sharing a single timebase (i.e., having been collected simultaneously) are packaged together. Sets of these “correlated sequences” are then packaged together as



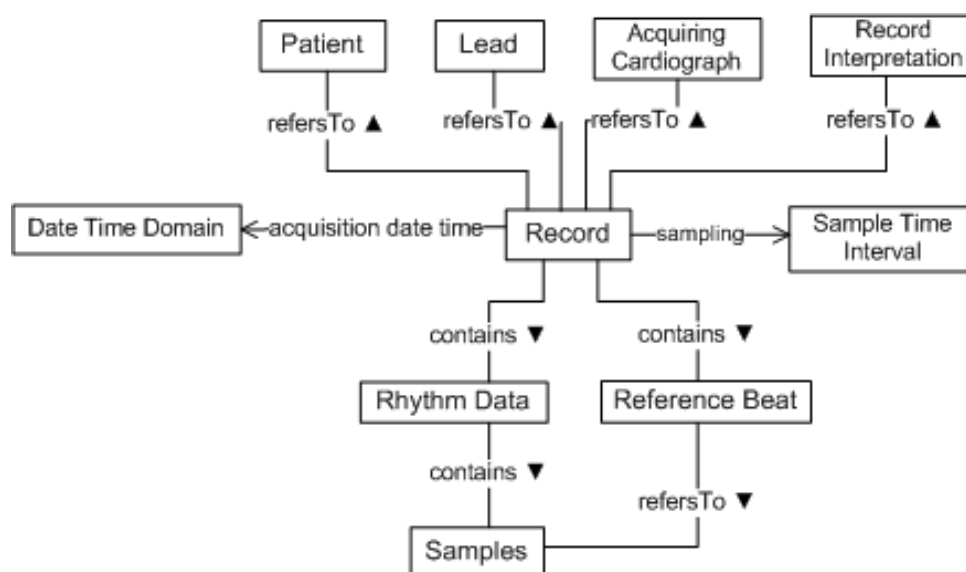


Figure 45: Conceptual model of the SCP-ECG.

a complete ECG session dataset. One can have a hierarchy of derived (filtered or otherwise transformed) representations based explicitly on some defined subset of the original data (regions of interest). Regions of interest also are used to bounder annotations. The resulting data format has nothing constraining its use to electrocardiographic data, or even time series data. What does produce this specialization is a domain-specific vocabulary for naming leads and types of annotations (and a convention for the units of measure).

Some aECG term textual definitions are listed below:

- *aECG* - Annotated ECG. The name given to a file or message conforming to HL7's "Annotated ECG" standard. It contains one or more series of ECG waveforms pertaining to a relative time point and a set of derived ECG findings for that time point.
- *Annotation* - An observation made on or associated with a series. E.g. a P-wave onset, a period of atrial fibrillation, the point at which a drug dosage was administered, etc.
- *Digital ECG* - A collection of digital information that contains ECG waveforms represented as sequences of numbers.
- *Digital Electrocardiograph* - A microprocessor-controlled electrocardiograph that captures the ECG waveforms using analog-to-digital converters and stores the waveforms as sequences of numbers. It produces digital ECGs.
- *Electrocardiograph* - A device that records the electrical activity of the patient's heart by tracing voltage-vs-time waveforms on paper.
- *Electrocardiogram* - Traditionally 12 waveforms (leads) arranged on a piece of paper representing 10 seconds of cardiac activity while the patient is lying on his back at rest. It is the physical record of the patient's cardiac activity produced by an electrocardiograph.
- *ECG* - Electrocardiogram. The term has also been used to mean any set of cardiac waveforms (leads) representing any contiguous period of time.

- *Lead* - A vector along which the heart's electrical activity is recorded as a waveform.
- *ROI* - Region Of Interest. Used to define a region within an ECG series so an annotation can be associated with it. E.g. the region of interest between the onset and offset of the P-wave can be associated with a P-wave annotation.
- *Series* - Contains one or more sequence sets sharing a common frame of reference. Because all sequence values are in the same frame of reference, the values are comparable. E.g. all relative time values are relative to the same point in time, and all voltages within Lead II are from the same pair of electrodes, and the subject is in the same "state".
- *Sequence Set* - A set of sequences all having the same length and containing related values. E.g. the 2nd value of a sequence is related to the 2nd value of every other sequence in the set.
- *Sequence* - An ordered list of values sharing a common code. E.g. sequence of voltage values with code "LEAD II".
- *Value* (p. 35) - The list of values in the sequence. The list of values can either be generated from a simple algorithm, or can be explicitly enumerated.
- *Subject* (p. 14) - Identifies the subject from which the ECG waveforms were obtained.
- *Series effective time* (p. 26) - Physiologically relevant time range assigned to the ECG waveforms contained within the series. This is typically referred to as the "acquisition time" which is determined by the device that collected the waveforms. If the device just recorded the beginning of acquisition, the time would go into the low part of the interval. If the device just recorded time when the data collection was finished, the time would go into the high part of the interval.
- *Annotated ECG component* (p. 26) - The component parts of the aECG. These are the waveforms and annotations. Even though the XML schema says this is an optional part of the aECG message, the message is not very useful without at least one series containing at least one waveform.
- *Series author* (p. 28) - This describes the device that "authored" (recorded) the series waveforms. This would typically describe an electrocardiograph or Holter recorder.
- *Annotation set component* (p. 39) - The annotation set is made up of one or more annotations (the components of the set).
- *Annotation* (p. 39) - An "annotation" is an observation made on the series by the annotation set's author. For example, if the electrocardiograph has algorithms to find the beginning of every QRS, an annotation set authored by the electrocardiograph could be made with component annotations for every QRS it finds. If the algorithm can also suggest a disease diagnoses (i.e. "interpretation"), the annotation set could include interpretation statements. If the algorithm can measure the heart rate, the measured rate can be included as an annotation.
- *supporting ROI* (p. 39) - Specifies where the observation was made within the series.

*Series* is an important element in the aECG data format which deserves a special attention. We then add some passages that refer to it. "A series contains all the sequences, regions of interest, and annotations sharing a common

frame of reference. [...] Typically a series will contain all the waveforms and annotations for a single ECG. If multiple ECGs are contained within a single aECG file, a different series is used for each. A series can be derived from another series. For example, a series containing representative beat waveforms is algorithmically derived from a rhythm series. Or, a series containing waveforms with special filtering applied can be algorithmically derived from the “raw” rhythm waveforms” (p. 26). [Obs: if two sequence sets were collected from different leads then they cannot be part of the same series, cf. (p. 26)]. A series can be yet of two different types (p. 26):

- *Rhythm* - the series contains rhythm waveforms. These are the waveforms collected by the device. The voltage samples are related to each other in real time (wall time).
- *Representative Beat* - the series contains the waveforms of a representative beat derived from a series of rhythm waveforms. The voltage samples are related to each other in time that’s relative to the beginning of the cardiac cycle, not real time.

From this textual description, we have assigned a conceptual structure to the AHA/MIT-BIH which is presented in Figure 46 (a data model) and Figure 47 (a conceptual model).

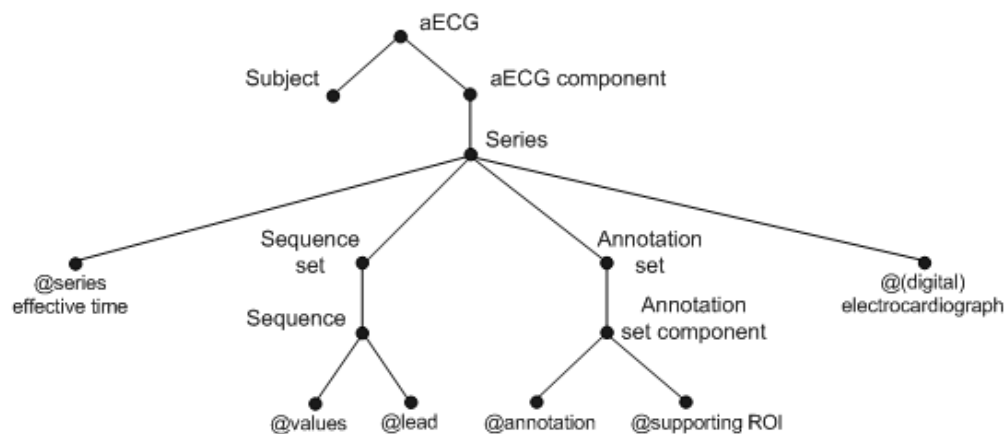


Figure 46: Tree-based data model of the FDA XML / HL7 aECG. The black circles denote nodes and the ‘@’ symbol indicates leaf elements

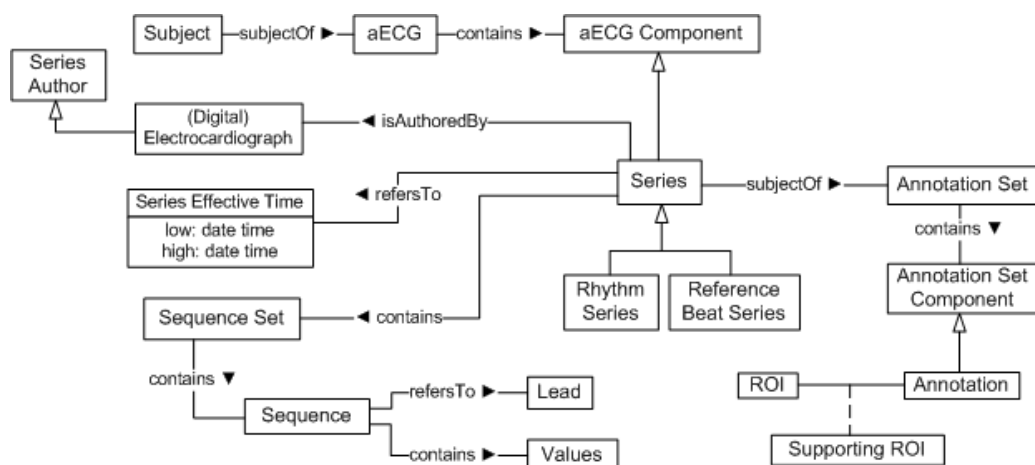


Figure 47: Conceptual model of the FDA XML / HL7 aECG.

## 7.2.4 Discussion

One can notice in the descriptions above that the referred elements mostly refer not to real entities in clinical reality, but to their information counterparts at the “symbolic-level”. By the latter we mean the symbolic world in which computer programs operate. That is evidenced by the use of words like “code”, “ID”, “header”, “section”, etc.

Besides, consider the following assertion in the aECG conceptual model (cf. Figure 47 on the left): a *(Digital) electrocardiograph* is-a *Series author*. An ontological analysis points out that this is a mistake since the former happens to be a substantial sortal (an instance of *kind* in UFO) and the latter is an instance of *role mixin*, i.e., an abstracted property which is anti-rigid and common to different roles. Intuitively, from the fact that an electrocardiograph can eventually lose the capability of being a series author it follows that it cannot be a type of series author - cf. the *is-a* definition prescribed in Table 1. Most likely, what was supposed to be said is that the *(Digital) electrocardiograph as a recorder* is-a *Series author*. This analysis is, of course, grounded in the real world, not in a symbolic one governed by laws somewhat different to those applying in the former. Overall, we reflect on this example to (i) emphasize the low quality of the conceptualizations underling the ECG data formats we have at hand; and also (ii) to demonstrate how the use of a low-expressive language with no support of ontological foundations may lead to bad modeling decisions.

Indeed, to move freely between a focus on what exists and one on the information entities that refer to what exists is not trivial. We therefore have been beset by this issue in the capture of the conceptual models introduced above. Nevertheless, it may be worth to say that our main point is not to propose ultimate conceptual models of those ECG standards, but rather to develop our ideas with respect to the use of domain reference ontologies, and the ECG Ontology in particular, to foster interoperability in Health Informatics.

As we have seen in the sections just presented, first of all, AHA/MIT-BIH represents the domain by focusing mainly in storage issues. SCP-ECG in turn address mostly communication issues, while the FDA XML / HL7 aECG favors flexibility towards an Internet-orientation, and presentation issues of the ECG waveform (though not observable here) as well. Altogether, as their requirements vary strongly, some heterogeneity in the data-level has in fact been expected; nonetheless, as they, on the other hand, deal with the same domain (the ECG), it would be intuitive to expect them to share (even a core) conceptualization. However, as we have seen, there is heterogeneity at the conceptual-level as well. In what follows, we make the case that the ECG Ontology can be used as a means for fostering a cost-effective integration among these ECG standards.

## 7.3 Ontology for Semantic Interoperability

The ECG standards treated here foster interoperability in Health Informatics by addressing one of the most applied exams in health environments. Nonetheless, if we go ahead into it to push the “interoperability envelope”, we might actually require the ECG standards to converge into a single one. Although much progress has been made just by the use of ECG standards (even heterogeneous to each other), Health Informatics would benefit even more from a unique “universal” ECG standard. Existing ECG data format standards like those presented here are intended to foster interoperability among their compliant information systems in their own right. Along these lines, if a (distributed) system follows one of these standards, their units (say, in a given hospital) are able to communicate data to each other. However, the problem still shows up if a, let us say, HL7-based system needs to communicate to a SCP-ECG-based one. As a matter of fact, converting schemata like the SCP-aECG converter (23) have been

developed in an attempt to overcome this issue. Time will tell us if such an alternative is able to be an ultimate solution immune to: (i) the *False-Agreement* problem (12), i.e., when the data is successfully exchanged but the underlying semantics between the two systems does not actually match; (ii) the increasing complexity in the ECG data representation resulting from analysis of advanced programs (e.g., some of those reported in *PhysioNet*, which outline new sorts of annotations). Overall, we report once more James Cimino's practical experience as putting effort for interoperability in Health Informatics (10, p. 394).

*"...the differences between the controlled vocabularies [or, in general, conceptualizations] of the two systems was found to be the major obstacle - even when both systems were created by the same developers."*

Indeed, if we consider the three somewhat different ECG domain conceptualizations that underlie the data formats just mentioned, one might rise the following question: what is the right one? And as each of them address its genuine purposes and requirements in its own right, another question then turns out: how to evaluate which purposes are more fair? How to make them converge? Naturally, this is only possible by relying on some shareable anchor. Something that, no matter what community alpha is in need, likely different to the need of community beta, could draw the attention of both to the fact that there exist acquiring devices, periodic observations carried out by them, samples resulting from these observations, geometric patterns of interest for diagnostic that emerge from those samples and so on. This anchor, as discussed in Section 3.3, has been over the centuries the object of interest of Philosophy and Science. A discipline grounded in tracking for referents in reality as it is the very business of Ontology, and particularly advocated by Barry Smith in (28), seems to be a step forward for supporting initiatives toward such "universal EHR". Along these lines, we can look *at* the instances existing not in medical terminologies or information systems, but in the health environment where the patient is subject of care by the physician.

Indeed, the ECG Ontology has been developed by employing this principle as far as we could. If we assume that the ECG Ontology does justice to *what the ECG is at the point of care* and solely this - i.e., regardless of technological issues that arise in representing it in a given information system; it could then be used to support the design of interoperable<sup>4</sup> versions of ECG data formats like AHA/MIT-BIH, SCP-ECG and FDA XML / HL7 aECG. By taking the ECG Ontology as a reference, the entities present in these data formats could be semantically mirrored to the ontology universals, instead of being object of pairwise mappings like SCP-ECG to HL7 aECG, vice-versa, and so on. Thereby, the ECG data formats should meet Cimino's desiderata (10), namely: (i) non-vagueness, the entities which form the nodes of the data format must correspond to at least one universal in the ontology; and (ii) they must correspond to no more than one universal, i.e., non-ambiguity. Since the ECG Ontology axiomatization allows little freedom to both vagueness and ambiguity, this solution would at least force the data formats to make their assumptions explicit. Besides, this proposal is cost-effective, since  $n$  data formats require  $n$  mappings to a reference ontology, whereas  $n(n-1)/2$  pairwise mappings would be required (98).

With that spirit, we report in the next section an integration experiment that supports the line of argument we have been defending here.

<sup>4</sup>Naturally, we are not referring to an interoperability procedure - as of messaging systems in the sense of computer networks, but to a structural disposition for exchanging information by sharing the same semantics.

## 7.4 An Integration Experiment

Once we have grasped the conceptual models underlying the ECG data formats presented in Section 7.2, we are able to try their integration. Along the lines discussed in the previous section, we conduct the following experiment. Each class of each ECG data format has been object of analysis in order to match a correspondent class of the ECG Ontology.

By starting with the AHA/MIT-BIH, we make use of its conceptual model depicted in Figure 43. The result is then presented in Figure 48. We then turn to the SCP-ECG conceptual model depicted in Figure 45 to find correspondence of their classes to the ECG Ontology universals, see Figure 49. Finally, the FDA XML / HL7 aECG conceptual model (Figure 47) is compared to the ECG Ontology as illustrated in Figure 50.

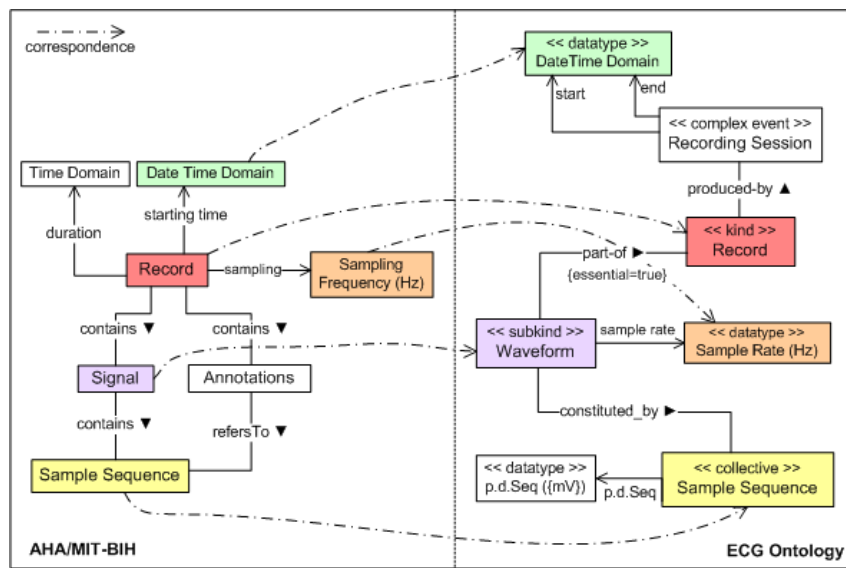


Figure 48: Integration between the AHA/MIT-BIH conceptual model and the ECG Ontology. The *starting time* and *duration* properties refer to *record* in AHA/MIT-BIH. This is odd since a record is a continuant, not an occurrent; but even then these properties loosely correspond to the recording session's start and end time in the ECG Ontology. The *Annotations* class misses a correspondent one in the ECG Ontology, since it would correspond not to the *quality* (or datatype in OntoUML) *Annotation* in the ECG Ontology, but rather to a multitude of annotations justifiable only as a programming resort.

Besides a bit of heterogeneity in term adoption, the symbolic orientation of the conceptual models of the ECG standards is the main source of more serious heterogeneity. In virtue of that, for instance, the important class *Annotations* misses a correspondent one in the ECG Ontology. This is because technological motivations such as efficient data access have led their maintainers to use a single file with few distinguishing fields to keep several annotations of very different sorts. On the other hand, the *Sequence* element in aECG exemplifies how an integration endeavor guided only by term alignment can lead to misleading associations. In this case the entity *Values* is actually that which corresponds to the fundamental *Sample Sequence* universal.

As result of the integration of each ECG standard to the ECG Ont., we have achieved an indirect matching between them as described in Table 9. Recall that we use *correspondence* here as a relationship that gives to the ECG standards' symbolic elements a real-world semantics according to the ECG Ontology. Ergo, this relation is not one of *equivalence* nor *identity*, i.e., two entities holding a correspondence relation are not the same.

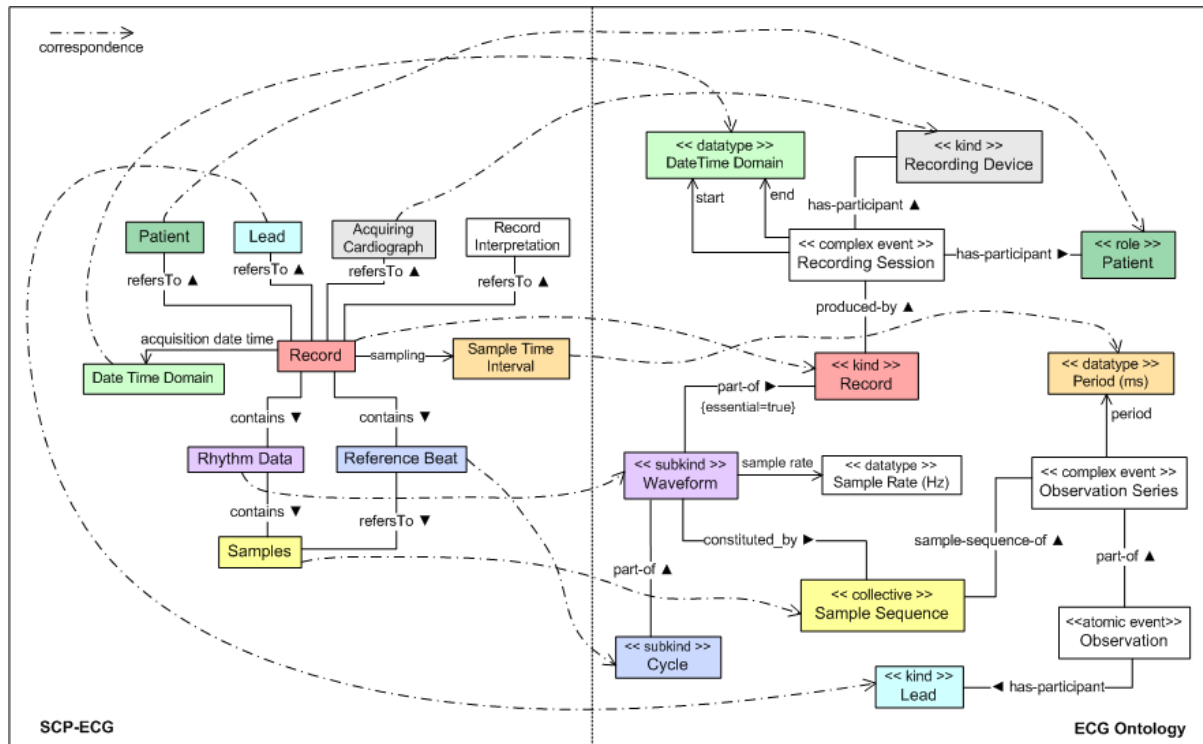


Figure 49: Integration between the SCP-ECG conceptual model and the ECG Ontology. In SCP-ECG, the time associated to a *record* (*acquisition date time*) has a ambiguous meaning; it could intuitively be either the start or the end date time. Nonetheless, it somehow corresponds to the start and end date time of the *recording session* in the ECG Ontology. Instead of referring to the record's *sample rate* (or *sample frequency*, in AHA/MIT-BIH), SCP-ECG refers to *sample time interval*. The latter happens to correspond to the *period* of the *observation series* that resulted in the *samples*, or *sample sequence* in the ECG Ontology. *Record interpretation* comprises an interpretation of the record as a whole, and has no correspondent entity in the ECG Ontology. It seems like such a general interpretation of the record interpretation could be obtained from the interpretation of elementary forms (viz., normal / abnormal), but it is not clear what sort of interpretation the SCP element precisely is.

Table 9: Correspondence relations between classes in the ECG Ont. and the ECG standards.

ECG Ontology	AHA/MIT-BIH	SCP-ECG	aECG
ecgOnto:Record	Record	Record	aECG
Sample rate (Hz)	Sampling frequency	-	-
Period (ms)	-	Sample time interval	-
ecgOnto:Sample sequence	Sample sequence	Samples	Values
ecgOnto:Waveform	Signal	Rhythm data	Rhythm series
ecgOnto:Cycle	-	Reference beat	Reference beat series
(Date time domain) - start and end time	(Date time domain) - starting time	(Date time domain) - acquisition date time	(Series effective time) - low and high
ecgOnto:Recording device	-	Acquiring cardiograph	-
ecgOnto:RD as recorder	-	-	(Digital) electrocardiograph
ecgOnto:Patient	-	Patient	Subject
ecgOnto:Lead	-	Lead	Lead
Annotation	-	-	Annotation

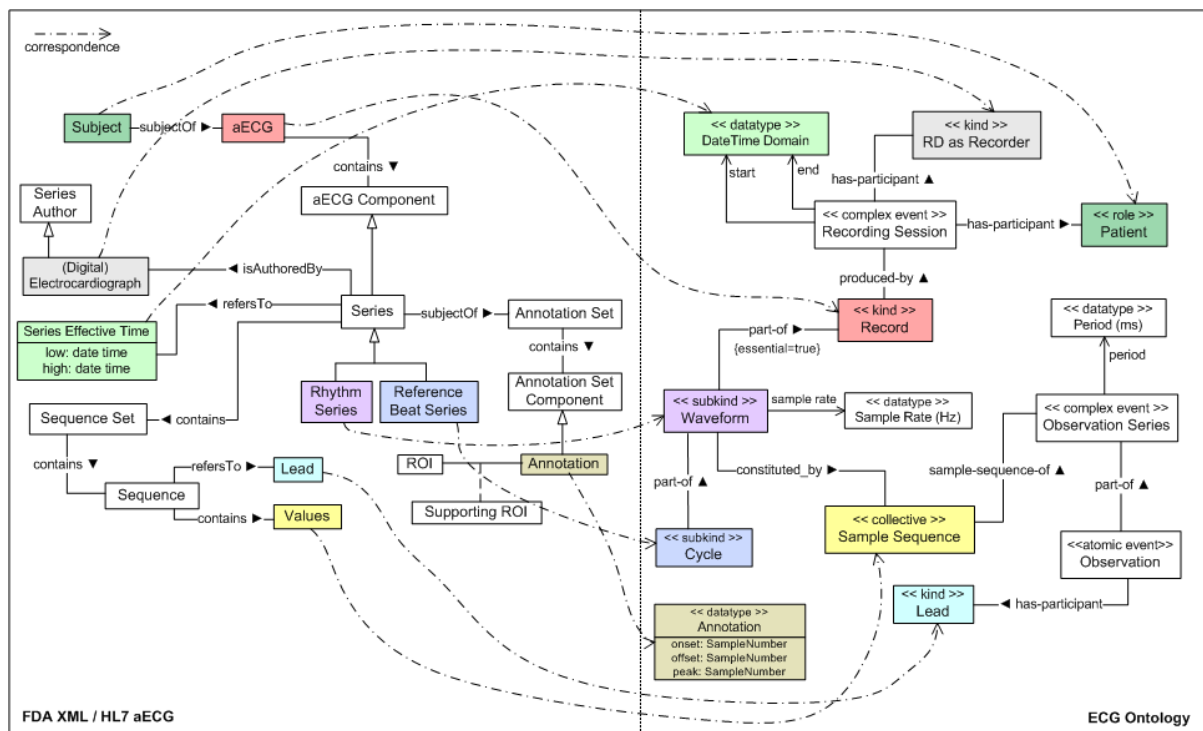


Figure 50: Integration between the FDA XML / HL7 aECG conceptual model and the ECG Ontology. *aECG component* is only meaningful in aECG, since there are many other aECG components related to organizational issues in this standard. *Series*, albeit the entity that encompasses all the ECG data in an aECG message (or *record*, in the ECG Ont.), does not correspond to any entity in the ECG Ont. It is actually relevant only in case the ECG waveform(s) is(are) fragmented into parts for addressing the structuring schema of ECG processing or viewer programs. The same holds for *Series author*, *Sequence*, *Sequence set*, *Annotation set* and *Annotation component*. The element *ROI* - Region of interest - approaches some of the ECG Ont. universals, but can hardly be assigned to anyone of them. It denotes any part of the ECG waveform that can be of some interest; and by its textual definition given in the aECG document it could be even the whole waveform. However, it lacks a fundamental property for that: it is not an *ECG form*, but it seems to be rather an arbitrary interval projected into the time axis. As we have not included interval entities in the ECG Ont., *ROI* comes to miss an associated ontology universal. *Supporting ROI* is then not able to be associated neither.

In sum, we submit that our integration experiment provides evidence for the following statement: *the ECG Ontology can be effectively used to foster interoperability between existing ECG standards.*

## 7.5 Conclusions

In this chapter we have applied the ECG Ontology to foster interoperability in Health Informatics. The key points worth to recall are:

- Even though the ECG standards address different purposes that could presumably lead to heterogeneity in their data models, their underlying (even core) conceptualizations of the ECG domain are also heterogeneous.
- A domain reference ontology can be used as a cost-effective means to foster interoperability of heterogeneous conceptual models.
- We have conducted an integration experiment which provides evidence that the ECG Ontology can be



effectively used to support the design of interoperable versions of those ECG standards.

In what follows we move from such an “off-line” application of the ECG Ontology to an “on-line” one in the field of symbolic AI.

## 8 Application in Symbolic AI

Consider, as an example, the cardiac electrical impulse (CEI) generated by the SA node of the heart of patient John Doe at a certain time in the past. If we have an ECG that “recorded” John’s heart activity at that moment, we can then look *at* that CEI by reconstructing it from John’s ECG. This is possible by, say, computing the rules presented in Section 5.6. Of course, the machinery developed in this thesis is still very basic since we can only “instantiate” such CEI under a normal/abnormal rubric. However, as long as we get further into the quagmire of elaborating the heart electrophysiology representation and then enrich the mapping relations introduced in this thesis, the properties of that virtual CEI could approach ever more the real CEI that took place possibly years ago in John’s heart. Links between geometric models of the heart anatomy and differential equations of myocytes’ depolarization could be established<sup>1</sup> as well for allowing advanced computer graphic simulations of such CEI. Could the EHR of the future store such a virtual CEI? Could this machinery reach a level of edification so powerful as capable to offer a germane support for physician’s reasoning towards a differential diagnosis?

Although the possibilities that unfold from the application of biomedical ontologies such as the ECG Ontology are promising, this chapter is devoted to introduce a less-pretentious application. We present here a reasoning-based web application that reconstruct electrophysiological phenomena mapped in a given ECG waveform. Our limit here is to reason whether or not a user-selected and faithfully annotated ECG elementary form maps a normal or abnormal electrophysiological phenomenon and which one it is. The user can then be informed about the result of such a deduction process. In case we have a normal phenomenon in hand, a flash animation takes place on an image of the heart conducting system to expose the user to the correlated CEI conduction process, see Figure 51. The animations developed are connected to and fired by a reasoning service built upon the ECG Ontology. The ECG data used as input is faithfully annotated and comes from *Physionet* (137).

We start the presentation of this application in Section 8.1 by introducing the ECG data used as input for the application. We then proceed in Section 8.2 to describe the technologies used and provide a schematic overview of the application’s architecture. In the following we elaborate on the application behavior in terms of its programming logic (Section 8.3). A performance evaluation is then reported in Section 8.4 with the main purpose of providing an account of how efficient are the reasoning services carried out over the ECG Ontology. In what follows (Section 8.5), we discuss how the application presented here fits in the terms of biomedical ontologies’ application; and refer to some correlated work in the field of educational animations for enabling us to say something about our own. At last, final considerations are summarized in Section 8.6.

---

<sup>1</sup>In the Virtual Soldier project (cf. Section 3.4) a similar logic has been applied to bridge heart anatomy geometric models to the heart anatomy modeled in the FMA.

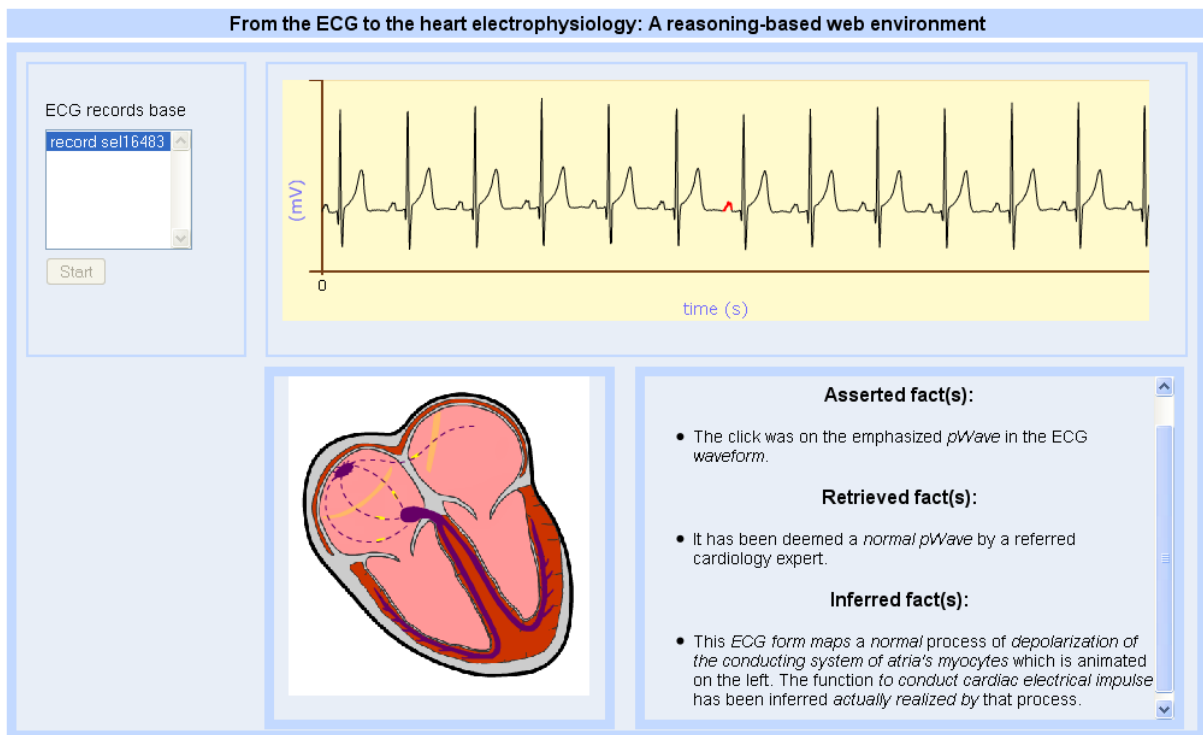


Figure 51: Screenshot of the reasoning-based web application.

## 8.1 ECG Data Input: The QT Database from Physionet

The ECG data used as input in this application is borrowed from the QT database (143). The QT Database is part of the Physiobank and contains a total of 105 fifteen-minute excerpts of two-lead ECGs. Within each record, between 30 and 100 representative beats were manually annotated by cardiologists, who identified the onset, peak and offset of the P-wave, the beginning and end of the QRS-complex (the QRS fiducial mark, typically at the R-wave peak, was given by an automated QRS detector), the peak and end of the T-wave. The criteria for beat selection were: (i) all are classified as “normal” by the system ARISTOTLE (144), and (ii) the preceding and following beats are also normal. Beats were only annotated during the final 5 minutes of the excerpts in order to allow analysis algorithms a minimum of 10 minutes for heuristic learning. In all, 3622 beats have been annotated by cardiologists. These annotations have been carefully audited to eliminate gross errors, although the precise placement of each annotation was left to the judgment of the expert annotators (143). All records were sampled at 250 Hz, which means an observation period (or sample interval) of 0.004 second. The records we use in the application presented here have been arbitrarily chosen from the QT database. We have populated the ECG OWL Ontology with these data.

## 8.2 Application Technologies and Architectural Overview

For handling the OWL ontology in memory, we have used the Java APIs Jena (145) and Pellet (133) just as reported in Section 6.3. The Jena API allows the building of OWL/RDF models and provide methods for accessing and modifying them. Jena is used also for providing to Pellet an access to the OWL ontology loaded in memory. Thus, the reasoning itself is fully supported by Pellet. The choice for Pellet is due to its interesting characteristics

reported in the literature. Pellet is efficient, customizable and can generate reasoning log information in detail. Moreover, it affords decidability even using SWRL rules - it applies the DL-safe rules' strategy (119); as well as integration (with consistency validation) between OWL restrictions and facts produced by SWRL rules.

The reasoning-based web application proposed here has been fully implemented in Java by using the GWT framework<sup>2</sup>. GWT affords a client-server architecture in which asynchronous remote procedure calls (RPC) take place. As illustrated in Figure 52, on the server side the main components are the ECG OWL ontology containing the ECG data, the Pellet reasoning engine, the Jena ontology model, and the ECG Chart Generator. Whereas on the client side they are the ECG chart and heart cond. system flash objects together with a widget of record samples to be selected; all they are included in the GWT Entry Point in the index page. Notice that both the ontologies and the flash objects are building blocks uncoupled from the application backbone.

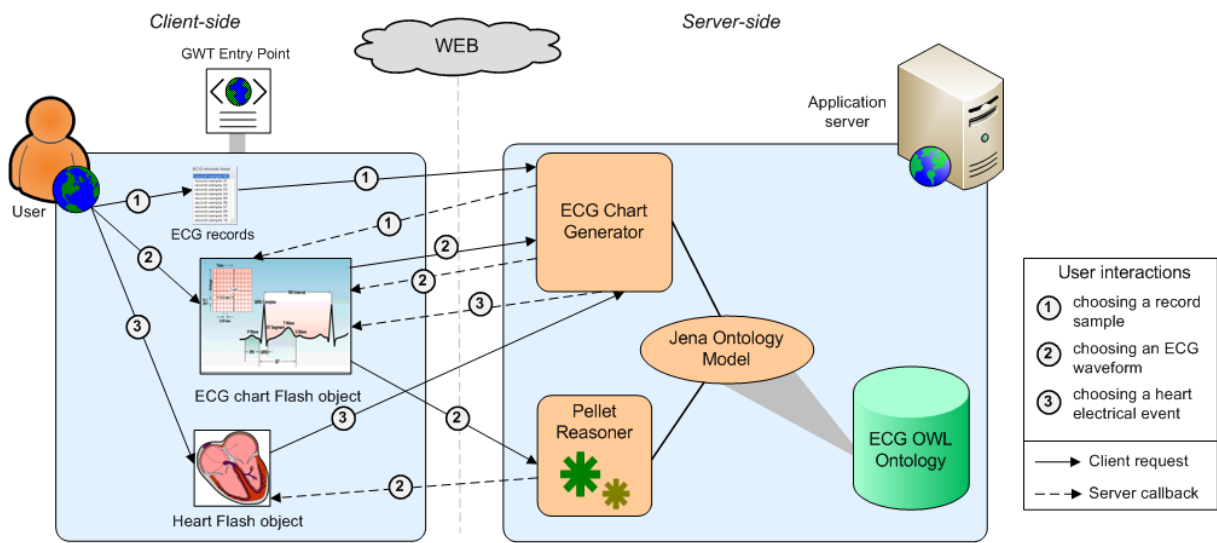


Figure 52: Application architectural overview.

## 8.3 Application Logic

The application exploits user-interaction by processing clicks either on an ECG chart or on the image whereby heart phenomena are illustrated via animations. A reasoning engine then draws (ontology-based) logical inferences as a result, by taking into account the mapping from the ECG to the heart electrophysiology. The facts either asserted (drawn by user input) or inferred (by reasoning results) are made explicit by means of (i) a short text written in a log box, (ii) the animation that simulates (an abstraction of) the heart conducting system, and (iii) an emphasized form on the ECG chart.

The application allows three basic user interactions (see Figure 51):

- **(I1):** selection (always by clicking) of an ECG record sample: this loads the record waveform from the ontology into the ECG chart. The waveform is clickable to make possible interaction I2 below.
- **(I2):** selection of a point on the ECG chart: this enables the reasoning engine to answer which ECG pattern is associated with this point. The reasoning result is then prompted in a log box (see Figure 51). Two cases are

<sup>2</sup>Google Web Toolkit. Available at: <<http://code.google.com/webtoolkit/>>.

possible, depending on whether or not the pattern fetched (the ECG elementary form) have a direct mapping to an electrophysiological process. In case it does, i.e., the clicked point is located in either the P wave, QRS complex or T wave, (i) a full message lets the user know the reasoning results for the whole mapping chain we are able to infer from that particular ECG elementary form. Besides, the heart conducting system image is animated to simulate the electrophysiological process mapped by it. Finally, the ECG chart is reloaded for emphasizing the recognized pattern. Otherwise, i.e., the clicked point is located in a segment, (ii) a simple message is shown in the log box and the pattern is emphasized in the ECG chart.

- **(I3):** selection of a specific region of the heart conducting system in the image. This enables (i) the animation of the bioelectric phenomenon that normally takes place on this region; and (ii) also reloads the ECG chart, but emphasizing the ECG pattern correlated to the phenomenon just mentioned in all the cycles appearing in the ECG waveform.

In what follows, that is described in the context of the two client-server RPCs they produce.

### 8.3.1 ECG Chart Service

When the user clicks a record sample (I1) (cf. the record samples' widget on the left in Figure 51), an RPC is triggered by the client requesting a URL. This URL locates a temporary file required by the chart flash object to plot the ECG waveform on the chart. This file contains the chart data generated by the server from the chosen record sample filled in the ontology. As soon as the client receives a successful RPC callback from the server, the chart flash object is loaded and the ECG data is populated on the chart.

On the other hand, when the user click comes from the heart conduction system flash object (I3), an extended version of the RPC just mentioned is triggered with an additional parameter. It informs the clicked region of the heart conduction system. The server then generates the correlated ECG chart, as mentioned above, with the ECG pattern associated with the clicked part of the heart conduction system emphasized in all cycles.

### 8.3.2 Inference Service

This service is requested by the client as a click is performed on the ECG waveform (I2). The service comprises an RPC passing as parameters the current selected record sample as well as the  $x$  coordinate of the chart clicked point. In the server-side logic, the following actions are triggered:

1. searching in the ECG ontology - where the record sample data is represented - which ECG pattern (elementary form) the clicked point is within;
2. reasoning about the fetched elementary form and infer new facts according to its properties (e.g. whether or not it maps an electrophysiological process, and whether or not the latter is normal);
3. enabling the animation that simulates the process if this is the case according to the reasoning results;
4. requesting the ECG chart service for reloading the chart with the recognized pattern emphasized.

For an example, consider the case captured by the snapshot shown in Figure 51. The click took place in the seventh cycle depicted in the waveform. As soon as the server has fetched the clicked pattern in the ontology and

recognized it as a *P wave* in that *cycle*, one fact indicating that that *P wave* instance has been clicked is asserted into the ontology model in memory. The reasoning concerning the selected elementary form is then performed. Since that particular *P wave* is normal, then it follows that it maps some normal process of depolarization of CSA myocytes. As *F4* is fired (cf. Subsection 6.2.5), the heart behavior associated to the elementary form in hand is inferred and described in the log box. Namely, it is entailed that the function *to conduct CEI* has been *actually realized* by the process mapped by that *P wave*. This process is then simulated through the flash animation and the ECG chart reloaded by showing the *P wave* instance emphasized (cf. Figure 51). Those correlations between ECG patterns and heart electrophysiological phenomena are therefore made explicit to the human observer.

### 8.3.3 Flash Media Objects

In short, the logical connection between the user clicks on the ECG chart to the animation events and vice-versa is controlled by the reasoning procedure. The connection and control are possible due to (i) JavaScript-based interfaces externally callable which are set on the flash objects at design time; and (ii) the accessibility of these interfaces from GWT client-side code based on the GWT JavaScript Native Interface (JSNI). The ECG chart and heart conduction system flash objects are codified as SWF files and have about 64 kb and 35 kb respectively. These sizes are quite short for download on the side of the client (this is one of the features of Flash), but quite useful for exposing the user to the inference results in a visual fashion.

## 8.4 Performance Evaluation

In this section, we report an experiment conducted in order to evaluate the performance of the proposed application. It has been carried out by using the application server Apache Tomcat 6.0.16 on a Microsoft Windows machine featuring AMD Turion 1.8 GHz and 1GB of main memory. The web application presented here is implemented in Java 1.6.0\_02 by using GWT 1.4, Jena-2.5.7 and Pellet-2.0-RC5.

The performance evaluation comprises each of the three user interactions (I1, I2 and I3) introduced above. We measured the time spent in each relevant event performed either on the server- or client-side. The first and last measurements take place always at the user click and at the program final answer, respectively. An additional RPC has been created for sending the client timing measurements to the server log. Network timing has not been considered since our interest is rather solely on the amount of time spent with the ontology-related tasks. Finally, for each interaction, we have measured the amount of time its internal events take for ten times as a means for reaching reliable mean, median and standard deviation values. Tables 10 and 11 below provide the time spent in each relevant event that takes place as interactions I1 and I3 are processed.

The amount of time spent by loading the OWL files, viz., about 3 seconds (mean), turns out to be the most significative. This is no surprise since IO operations are more costly, but also because the construction of a not so short RDF graph as in the case of the ECG OWL Ontology in fact takes some time. The timing measurements for I2, however, are the foremost in relevance. The reason is that they comprise a timing parameter for automated reasoning over the ECG OWL Ontology and retrieving information from it. As shown in Table 12, these events are very fast in comparison to other processing tasks required in this application. Those timing measurements make up a strong evidence for the efficiency the ECG OWL Ontology is able to afford. This, therefore, comes to meet one of our objectives reported in Chapter 4.

Table 10: Timing measurements (in ms) for loading the chart with the ECG waveform plotted (I1).

Event	Mean	Median	Standard Deviation
Click processing	5	5	2
RPC call	66	61	24
Loading OWL files	3092	2875	631
Generating ECG flash data	78	40	78
RPC callback	23	20	13
Callback processing	2	2	1
Total response time	3266	3017	675

Table 11: Timing measurements (in ms) for reloading the ECG chart by emphasizing an ECG pattern in response to a click in the heart conducting system (I3).

Event	Mean	Median	Standard Deviation
Click processing	4	4	1
RPC call	46	43	27
Loading OWL files	3061	2891	587
Generating ECG flash data	97	47	85
RPC callback	20	18	9
Callback processing	6	6	1
Total response time	3234	3022	628

Overall, it is worth to mention that this application has been developed not striving for optimal performance, but rather for supporting some conjectures conveyed in the course of this thesis. We then outline, now also upon an empirical basis, that:

*the ECG OWL Ontology is an effective material for automated reasoning over ECG universals and particulars.*

Nevertheless, many optimization issues can be addressed in view of a better computational performance. Among them, we include: (i) load the OWL files only once during the whole application lifetime; for the moment they are loaded for each user interaction and closed after response; and (ii) use a more efficient chart library for exhibiting the ECG waveform; the Open Flash Chart library<sup>3</sup> which has been used in the current application version is, though visually appealing, more time consuming than we expected at a first glance.

## 8.5 Discussion

### 8.5.1 Reasoning over Universals and Particulars

At this point, it should be clear that the ECG Ontology is not meant for supporting pattern recognition in ECG data. Let us say, for the automated recognition of elementary forms in a given raw ECG waveform. Indeed, the ECG pattern recognition literature has shown better results with heuristic techniques - e.g., (144, 146), instead of symbolic reasoning. For this reason we fuel our application with annotated ECG data taken from Physionet. As

<sup>3</sup><http://teethgrinder.co.uk/open-flash-chart/>.

Table 12: Timing measurements (in ms) for getting from the ECG ontology information resulting from reasoning (I2).

Event	Mean	Median	Standard Deviation
Click processing	3	3	1
RPC call	8	11	6
Loading OWL files	3157	3006	678
Fetching ECG pattern	3	0	10
Reasoning	53	0	162
Information retrieval	345	282	242
RPC callback	24	16	11
Callback processing	2	2	1
Total response time	3594	3452	657

mentioned in Section 7.1, Physionet is one of the most referred sources for annotated ECG data resulting from ECG pattern matching systems, but also from physicians' analysis (147).

The ECG Ontology, on the other hand, is valuable for a semantically enhanced representation of ECG data enabling further inference. It bears a canonical model of heart anatomy and a canonical model of heart electrophysiology. The ECG model, contrarily, can be filled in by any real ECG record instance. However, a deformed QRS complex (possibly indicating some pathology) would not have a non-canonical cardiac electrical impulse to map to. Given this elucidation, the application just presented shed light on what can be done. By using an instance of a normal ECG record (an artifact for study), we reconstruct the (canonical) electrophysiology behind it. So, from a normal instance of QRS complex (faithfully annotated), we are able to reconstruct the cardiac electrical impulse behind it and the anatomy on which it has taken place.

All this could be done with a non-canonical ECG record as well if we had a non-canonical model of physiology to reconstruct. As far as we have investigated, that seems to be possible by extending the sub-ontology of heart electrophysiology to address fuzziness (vagueness) in the actual realization of heart electrophysiological functions. Our application, however, exploits the ECG Ontology in what it is currently able to offer. This application turns out to be useful, we shall discuss as follows, to support learning in Electrocardiography and heart electrophysiology.

### 8.5.2 Educational Animations

Computational technologies have been increasingly explored to make biomedical knowledge and data more accessible for human understanding, comparison, analysis and communication. One of the main supporting resources in this sense is the simulation of biomedical phenomena in a comprehensible visualization toolbox. According to the motto "*an image says more than a thousand words*", Wünsche points out that "*[v]isualization is an attempt to simplify those tasks*" (148), which in fact are non-trivial if we get deeper and deeper into the understanding of the mechanisms of the human body.

Along these lines, a commonsense logic-based representation of biomedical phenomena as they are scientifically explained in medical textbooks can be combined to the usual mathematical models to support human comprehensibility. Use scenarios range from aid learning in medical sciences to support physicians' decision



making. In the former scenario, students could be supported not only by visual media, but also by asserted and inferred facts about the biomedical data exhibited. Such features could (arguably) ease the recognition of relationships between visual patterns and what is behind them.

Indeed, in a similar fashion García et al. present several experiments that evidence benefits achieved in using flash animations to aid learning of Descriptive Geometry (149). We believe that flash animations could offer support for exploratory learning in Electrocardiography / heart electrophysiology as well. By interacting with such web media, students could actively explore the ECG and the heart conducting system in a goal-oriented constructive process. They could then actually visualize the immaterial electrical currents generated by the heart pacemaker cells. At a first glance, as presented above, we have chosen flash to present both the ECG chart and the heart conducting system animations. These are controlled by the inferences fired by the reasoning engine according to the ECG Ontology.

In sum, this is the usage scenario we feel to be of the most direct applicability for our reasoning-based web application at the moment. Nonetheless, the application is a prototype that still requires some effort to be employed w.r.t. optimization issues and experts' evaluation in order to be ready for real scenarios. As long as a release version is deployed, some pedagogic methodology - perhaps similar to what is adopted by Garía et al. in (149) - could be employed to exploit the application usability.

## 8.6 Conclusions

In this chapter we have applied the ECG Ontology in a reasoning-based web application. This application can be accessed and tried at our project website<sup>4</sup>. A previous version of it - which uses a previous version of the ECG Ontology as well - is reported in (150). Overall, the contents of this chapter can be summarized as follows.

- We have demonstrated by experiment that the ECG OWL Ontology is an effective material for automated reasoning over ECG universals and particulars.
- The reasoning-based web application presented here is a prototype that can illustrate benefits of using the ECG OWL Ontology for representation, reasoning and visualization of heart electrophysiology on the web.
- Although still a prototype, we claim this application to be potentially useful to offer support for interactive learning in Electrocardiography / heart electrophysiology.

This thesis has then reached its final considerations and remarks that are the object of the next and last chapter.

---

<sup>4</sup><http://nemo.inf.ufes.br/biomedicine/ecg.html>.

## 9 Discussion & Final Considerations

This chapter provides a discussion of the contribution and significance of this thesis. We touch upon limitations of the work presented here as well. The chapter starts by revisiting the thesis' research questions and concludes after posing open problems to be addressed in further work.

### 9.1 Revisiting our Goals and Research Questions

The main goal of this thesis has been: *“to develop an ontological theory of the ECG (independent of application and codification language), and further apply it by providing evidence of its benefits”*.

This goal has been fulfilled by the following results. First, by an ontology of ECG, and second, by its twofold application (i) to foster interoperability of ECG standards and (ii) to reason over ECG universals and particulars in a web application. More specifically, the following specific goals are accomplished:

- Goal 1: We aim to develop two ontology artifacts: one ontologically well-founded theory of the subject domain meant to be strongly axiomatized for constraining as much as possible the theory's intended meaning; and another meant to be a computable artifact for automated reasoning and information retrieval.
- Goal 2: Provide evidence for the following hypothesis: an ECG reference ontology can be used to foster interoperability of different conceptual models in the ECG domain.
- Goal 3: Likewise, provide evidence for the assumption that an ECG ontology implementation derived from its reference counterpart can be used with genuine benefits in a reasoning-based computer application.

We argue that, as expected, these goals have been met in the course of this thesis according to the following. Goal 1 has been reached along Chapters 5 and 6. Chapter 7 made the case that Goal 2 could in fact be pursued and be satisfactorily evidenced. Finally, Chapters 6 and 8 put together Goal 3 into an effective empirical basis.

If this is the case, let us then put back here our research questions introduced in Section 1.2 and answer them by summarizing what has been developed throughout this thesis.

- RQ 1: *What is the ECG in essence?* The ECG is a cardiological exam which has been modeled here - in Chapter 5 - in depth under a principled ontological analysis. As a result it has been outlined an ECG ontology capable of answering fundamental questions regarding the ECG.
- RQ 2: *What can an off-line ECG ontological theory (or reference conceptual model) be used for?* As we have demonstrated in Chapter 7, one of the potential off-line applications of such theory is to support the

design of interoperable versions of ECG conceptual models in Health Informatics. Moreover, an ontological theory can support the building of such models with a special feature: be grounded not in specific (and perhaps ephemeral) interests, but in reality.

- RQ 3: *Is it worthwhile to derive an ontology implementation from an ontological theory?* By practically balancing benefits and drawbacks related to the tradeoff between expressiveness and computational tractability, we can derive a computable artifact from such an ontological theory (or reference ontology). As evidenced in Chapter 6, this ontology implementation cannot keep all the expressivity of the original reference ontology on which it is based. However, as evidenced in Chapters 6 and 8, the ontology implementation is worthwhile to be derived from it in virtue of inheriting some of its germane features.
- RQ 4: *What can be done by using the codification of an ECG ontology in a reasoning-based computer application? Are there any benefits, say, when compared to other AI formalisms? Which are them?* By relying on the reasoning features of a computable ontology of ECG we can reason over universals and particulars of this domain. This has been demonstrated in the reasoning-based application presented in Chapter 8. Besides, as long as research question RQ3 above is assumed to be answered positively, it follows that there has been genuine benefits in having derived such a computable ECG ontology from a sound ECG ontological theory.

In virtue of accomplishing the goals mentioned above and answering those research questions we believe are contributing to the biomedical ontology literature in particular, but also to the ontology engineering literature. The contribution for the former lies in the accomplishment of the goals exposed above and also by answering RQ1 with an underlying work to support it. The contribution for the latter relies also in meeting those goals as they can be seen as case studies for more general statements, but mostly by answering RQ2, RQ3, and RQ4.

## 9.2 Significance

For the past years, we have been dealing with the ECG as a subject of ontological inquiry. An initial effort of representing ECG data by applying Formal Ontology techniques resulted in a preliminary ECG domain ontology reported in (151, 152). Upon that preliminary work, we have built an early version of the reasoning-based application presented here which is reported in (153, 150). Since then, we have been revising the basis underlying that early endeavor. This has led us to reformulate our ECG ontological representation, for the sake of increasing specialization, degree of detail, density and connectivity, to cite the terms conveyed by (95, p. 335). The first publication reporting this second step in our iterative approach is (130).

The applicability of the ECG Ontology has been the essence of most of the questions raised in the presentations (especially in the first ones) of the papers mentioned above. Well, we hope Chapters 7 and 8 have provided evidence enough for the ECG Ontology applicability. Notwithstanding, we would like to rather highlight here something else, viz., that the ECG ontological theory *per se* can be even a contribution in terms of scientific development.

Along these lines, it is fair enough to say that the ECG Ontology is in its own right a resource meant for enabling a better (objective, logical) understanding of this cardiological exam. In this way, it can be (shared and) accessed by cardiology communities as a means for an evolving scientific-based common-sense representation of the exam. Instead of being tacit in the mind of experts, or even implicit in natural language-based assertions in

medical textbooks; such a logical representation could be a more fit and sharable object for the proper acception / refutation process that enables scientific development.

Altogether, as part of an ongoing worldwide research effort to foster ontological representations of biomedical reality, our endeavor is in place. Naturally, our ECG ontological inquiry may be elaborated to increase, say, the degree of detail, and even to cover eventual lacunae. Meanwhile, the challenge of ontology integration is still tough even in this ever more anchored research field of so-called biomedical ontology. However, by striving for keeping compliance with correlated initiatives, we have put an effort forward in this direction. Anyhow, it is the case that “[t]he value of any kind of data [or ontology] is greatly enhanced when it exists in a form that allows it to be integrated with other data [or ontology]” (17). In that spirit, the ECG Ontology can be understood as a contribution to be aggregated into the biomedical ontology effort.

## 9.3 Limitations

We identify in our work three main limitations: (i) first, we are restricted to a canonical heart electrophysiology; and (ii) second, we have covered only a single lead-ECG scope in our ontological theory; and (iii) third, that the implementation of the ECG Ontology does not take account of time.

The first issue is mostly due to the complexity in dealing with physiological aspects of the human heart. This is particularly tough when both genotypic and phenotypic issues are to be covered. Therefore, a strong research effort is required to extend the ECG ontological theory presented here with such a purpose. Nonetheless, if we take the FMA as an example, it is restricted to canonical anatomy but still has many application scenarios (much of them already *in use*) as evidenced by the literature, cf. Section 3.4. Along these lines, the application proposed in Chapter 8 supports by the same token the usability of our work even with such a limitation.

The second issue in turn concerns the extension of the ECG ontological theory to cover multiple lead ECGs. If compared to the former issue, this one is much less demanding in terms of effort. It can be covered by further specifying at least the twelve standardized leads (e.g., Lead II, aVR, V5, etc) as subtypes of the universal **Lead**, cf. Section 5.4. Besides, the subtypes of **Body surface region** on which an electrode can be placed to come up with each of those leads (e.g., the wrist of the right hand) can be specified as well. This seems to be enough for connecting these body surface regions to the correspondent leads by means of different *relators* and then extending the ECG Ontology to a 12-lead version.

Finally, the third issue is in virtue of a bit less scientific, and more practical challenge. Namely, that our schedule has not let us to get further into a suitable ontology codification framework for taking account of time. Such formalisms and accompanying machinery do exist and are reported in the literature. Some examples are (i) the early work developed by James Allen regarding an interval-based temporal logic proposed together with a computationally effective reasoning algorithm (154); or even more recent resources such as the OWL-Time ontology, which is meant for describing the temporal content of web pages and the temporal properties of Web Services (155). This ontology falls into the OWL-DL rubric, more specifically into the DL family SHIOF(D). It derives from the former DAML-Time, and is currently a W3C recommendation (156).

## 9.4 Open Problems and Future Work

As exposed in the discussion above, the limitations of our results deserve some research effort to get overcome. In particular, with respect to the required modeling elaboration of the heart electrophysiology, we have been inclined to think towards the following. It seems that a fuzzy account of the realization of heart electrophysiological functions could enable a suitable ontological modeling of heart disfunctions. As far as we have foreseen, this could make possible to capture how distortions in the ECG elementary forms would impinge in the degree of realization of such functions. We believe this to be an important starting point to cope with particular pathological cases. At best, a (geometric) mapping between the ECG forms and such a degree of realization would enable an objective understanding not only of the impact of diseases as we apprehend them in scientific-based common-sense, but also of the interrelationships between them. If the substantial amount of work required is put in practice, then promising results seem to be reachable.

Among the envisaged directions for future work we include:

- for a short- and mid-term research and development: the extension of the ECG Ontology implementation to take account of time and measure the impact on the reasoning performance.
- for a short-term research and modeling effort: to extend the ECG ontological theory and then further its implementation to cover multiple lead ECGs (e.g., the 12-lead ECG).
- for a longer-term research: the extension of both the heart electrophysiology model to cover physiological disfunction and the ECG model to cover pathological issues as well.

## 9.5 Final Considerations

In this thesis we provide an ontological account of the cardiological exam ECG and its correlation to the human heart electrophysiology and anatomy. The ECG Ontology outlined here constitutes an axiomatized domain theory grounded in a principled ontological basis. The applicability of this ontology has also been enlightened for two different purposes, viz., managing heterogeneity of ECG data format standards and automated reasoning over ECG universals and particulars. With the latter in mind, we have (loosely) translated the models and FOL formulae we present here into the ontology codification language OWL DL with its SWRL extension.

Geometric models for anatomy and differential equations for physiology have been extensively used to simulate biomedical phenomena. Notwithstanding, we claim that a common-sense representation of these phenomena as they are scientifically explained in medical textbooks owes its *raison d'être* as well. They can be used to support the human user as he/she apprehend and reason about these phenomena. The ECG Ontology finds its place in this enterprise.

Altogether, the business of biomedical ontology is a prominent discipline in Medical- and Bio-informatics nowadays, and the results preliminarily reached point towards (gradually) filling the gap between basic biological research and medical applications. As nicely put by Yu in (26, p. 252), while achieving this would let biological researchers to benefit from harnessing biomedical representations that are increasingly stored in computable forms, it would further be a significant step towards fulfilling the vision that Blois described in 1988 (157):

The medical practitioner needs to be able to harness the tools of reasoning better to apply them to a mixture of low-, middle-, and high-level data. This is essential if physicians are to range back and forth, consciously and effectively, from the mathematical descriptions of atomic and molecular events to the statistical associations exhibited by complex biologic systems, and to the natural-language descriptions at the clinical and behavioral levels.

If this outlook sounds quite exotic or even too ambitious, it may be furthermore interesting to report a quotation from Drew McDermott (158) nicely cit. by Guarino in (159).

Those were the good old days. I remember them well. Naive Physics. Ontology for Liquids. Commonsense Summer. [...] Wouldn't it be neat if we could write down everything people know in a formal language? Damn it, let's give a shot! [...] If we want to be able to represent anything, then we get further and further from the practicalities of frame organization, and deeper and deeper into the quagmire of logic and philosophy.

In this thesis, we stand for Guarino's belief that "*this quagmire is well worthwhile getting into*" (159).

## References

- 1 MAINTAINERS. UMLS - Unified Medical Language System. Release November 2008. Project website: <<http://www.nlm.nih.gov/research/umls/>>.
- 2 MAINTAINERS. NCI Thesaurus. Release February 2009 (09.02d). Project website: <<http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do>>.
- 3 MAINTAINERS. Gene Ontology. Project website: <<http://www.geneontology.org/>>.
- 4 BUREK, P. et al. A top-level ontology of functions and its application in the Open Biomedical Ontologies. *Bioinformatics (Oxford)*, v. 22, n. 14, p. e66–e73, 2006. [doi: 10.1093/bioinformatics/btl266].
- 5 BLAKE, J. Bio-ontologies - fast and furious. *Nature Biotechnology*, v. 22, n. 6, p. 773–74, 2004. [doi: 10.1038/nbt0604-773].
- 6 KUMAR, A.; YIP, Y. L.; SMITH, B.; GRENONA, P. Bridging the gap between Medical and Bioinformatics: An ontological case study in colon carcinoma. *Computers in Biology and Medicine*, v. 36, n. 7, p. 694–711, 2006. [doi: 10.1016/j.compbimed.2005.07.001].
- 7 ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: The Foundational Model of Anatomy. *J. of Biomedical Informatics*, v. 36, n. 2003, p. 478–500, 2003. [doi: 10.1016/j.jbi.2003.11.007].
- 8 SCHULZ, S.; HAHN, U. Towards the ontological foundations of symbolic biological theories. *Artificial Intelligence in Medicine*, v. 39, n. 3, p. 237–250, 2007. [doi: 10.1016/j.artmed.2006.12.001].
- 9 HOEHNDORF, R.; LOEBE, F.; KELSO, J.; HERRE, H. Representing default knowledge in biomedical ontologies: Application to the integration of anatomy and phenotype ontologies. *BMC Bioinformatics*, v. 8, n. 377, 2007. [doi: 10.1186/1471-2105-8-377].
- 10 CIMINO, J. J. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine*, v. 37, n. 4-5, p. 394–403, 1998.
- 11 BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, p. 1–7, May 2001. Available at: <<http://www.sciam.com/article.cfm?id=the-semantic-web>>.
- 12 GUARINO, N. Formal Ontology and information systems. In: GUARINO, N. (Ed.). *Proceedings of the 1st Formal Ontology and Information Systems*. Amsterdam: IOS Press, 1998. p. 3–15. Trento, Italy.
- 13 BODENREIDER, O.; STEVENS, R. Bio-ontologies: Current trends and future directions. *Briefings in Bioinformatics (Oxford)*, v. 7, n. 3, p. 256–274, 2006. [doi: 10.1093/bib/bbl027].
- 14 ROSSE, C. et al. A strategy for improving and integrating biomedical ontologies. In: FRIEDMAN, C. P. (Ed.). *AMIA 2005 Annual Symposium Proceedings*. Washington, USA: [s.n.], 2005. p. 639–643.
- 15 SMITH, B.; KUMAR, A.; CEUSTERS, W.; ROSSE, C. On carcinomas and other pathological entities. *Comparative and Functional Genomics*, v. 6, n. 7-8, p. 379–387, 2005. [doi: 10.1002/cfg.497].
- 16 BITTNER, T.; DONNELLY, M. Logical properties of foundational relations in bio-ontologies. *Artificial Intelligence in Medicine*, v. 39, n. 3, p. 197–216, 2007. [doi: 10.1016/j.artmed.2006.12.005].
- 17 SMITH, B. et al. The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, v. 25, n. 11, p. 1251–1255, 2007. [doi: 10.1038/nbt1346].
- 18 ASHBURNER, M. et al. Gene Ontology: Tool for the unification of Biology. *Nature Genetics*, v. 25, p. 25–29, 2000. [doi: 10.1038/75556].

- 19 BROOKSBANK, C.; CAMERON, G.; THORNTON, J. The European Bioinformatics Institute's data resources: Towards systems biology. *Nucleic Acids Res.*, v. 33, n. Database Issue, p. D46–D53, 2005. [doi: 10.1093/nar/gki026].
- 20 RUBIN, D. L. et al. Ontology-based representation of simulation models of physiology. In: OHNO-MACHADO, L. (Ed.). *AMIA 2006 Annual Symposium Proceedings*. Washington, USA: [s.n.], 2006. p. 664–668.
- 21 COOK, D. L.; MEJINO, C. R. J. L. V. Evolution of a Foundational Model of Physiology: Symbolic representation for functional bioinformatics. In: FIESCHI, M. et al. (Ed.). *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO'04)*. Amsterdam: IOS Press, 2004. v. 107 Stud Health Technol Inform, n. Pt 1, p. 336–340.
- 22 GESELOWITZ, D. On the theory of the electrocardiogram. *Proceedings of the IEEE*, v. 77, n. 6, p. 857–876, 1989. [doi: 10.1109/5.29327].
- 23 MAINTAINERS. SCP-ECG - Standard Communications Protocol for Computer-Assisted Electrocardiography. Project website: <<http://www.openecg.net/>>.
- 24 MAINTAINERS. FDADF - FDA XML Data Format Design Specification. Available at: <<http://xml.coverpages.org/FDA-EGC-XMLDataFormat-C.pdf>>. Access on March 2009.
- 25 MAINTAINERS. HL7 ECG Annotation Message v3. Project website: <<http://www.hl7.org/V3AnnECG>>.
- 26 YU, A. Methods in Biomedical Ontology. *Journal of Biomedical Informatics*, v. 39, n. 3, p. 252–266, 2006. [doi: 10.1016/j.jbi.2005.11.006].
- 27 JOHANSSON, I. Bioinformatics and biological reality. *Journal of Biomedical Informatics*, v. 39, n. 3, p. 274–287, 2006. [doi: 10.1016/j.jbi.2005.08.005].
- 28 SMITH, B. From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. *Journal of Biomedical Informatics*, v. 39, n. 3, p. 288–298, 2006. [doi: 10.1016/j.jbi.2005.09.005].
- 29 PANFILOV, A. V.; HOLDEN, A. V. *Computational biology of the heart*. 1st. ed. [S.l.]: Wiley, 1997.
- 30 HAYES, P. J. The naive physics manifesto. In: MICHIE, D. (Ed.). *Expert Systems in the Micro-Electronic Age*. Edinburgh: University Press, 1978. cap. 4, p. 242–70.
- 31 SMITH, B.; WELTY, C. Ontology: Towards a new synthesis. In: SMITH, B.; WELTY, C. (Ed.). *Proc. of the 2nd International Conf. on Formal ontology in information systems*. New York: ACM Press, 2001. p. 3–9.
- 32 SMITH, B. Ontology. In: FLORIDI, L. (Ed.). *Blackwell Guide to the Philosophy of Computing and Information*. [S.l.]: Wiley-Blackwell, 2003. cap. 11, p. 155–166.
- 33 GUIZZARDI, G. On Ontology, ontologies, conceptualizations, modeling languages, and (meta)models. In: VASILECAS, O. et al. (Ed.). *Databases and Information Systems IV - Selected Papers from the 7th International Baltic Conf. (DB&IS'2006)*. Amsterdam: IOS Press, 2007. (Frontiers in Artificial Intelligence and Applications, v. 155), p. 18–39.
- 34 GUIZZARDI, G. *Ontological foundations for structural conceptual models*. PhD Thesis — University of Twente, The Netherlands, 2005. Available at: <<http://purl.org/utwente/50826>>.
- 35 SOWA, J. F. *Knowledge representation: Logical, philosophical, and computational foundations*. [S.l.]: Belmont, CA, USA: Brooks-Cole, 2000.
- 36 MEALY, G. H. Another look at data. In: *Proc. of the Fall Joint Computer Conference*. London: Academic Press, 1967. (AFIPS Conference Proceedings, v. 31), p. 525–534. Anaheim, USA.
- 37 QUINE, W. V. On what there is. In: QUINE, W. V. (Ed.). *From a logical point of view: Nine logico-philosophical essays*. Second revised edition. [S.l.]: Harvard University Press, 1953. cap. I.
- 38 HAYES, P. J. Naive physics I: ontology for liquids. Morgan Kaufmann Publishers Inc., San Francisco, USA, p. 484–502, 1990.



- 39 SOWA, J. F. *Conceptual structures: Information processing in mind and machine*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1984.
- 40 STEFIK, M.; CONWAY, L. Towards the principled engineering of knowledge. *AI Magazine*, v. 3, n. 3, p. 4–16, 1982. Available at: <<http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/374>>.
- 41 GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum Comput Stud*, v. 43, n. 5-6, p. 907–28, 1995. [doi: 10.1006/ijhc.1995.1081].
- 42 ARMSTRONG, D. M. *Universals: An opinionated introduction*. Boulder, Australia: Westview Press, 1989.
- 43 GUARINO, N.; WELTY, C. Identity, unity, and individuality: Towards a formal toolkit for ontological analysis. In: HORN, W. (Ed.). *Proceedings of ECAI-2000: The European Conference on Artificial Intelligence*. Amsterdam: IOS Press, 2000.
- 44 GUARINO, N.; WELTY, C. Evaluating ontological decisions with ONTOCLEAN. *Communications of the ACM*, v. 45, n. 2, p. 61–65, February 2002. [doi: 10.1145/503124.503150].
- 45 MAINTAINERS. OntoClean. Project website: <<http://www.ontoclean.org>>.
- 46 WELTY, C.; GUARINO, N. Supporting ontological analysis of taxonomic relationships. *J. Data and Knowledge Engineering*, v. 39, n. 1, p. 51–74, 2001.
- 47 SOWA, J. F. Sowa's Ontology. Project website: <<http://www.jfsowa.com/ontology/>>.
- 48 MAINTAINERS. DOLCE - Descriptive Ontology for Linguistic and Cognitive Engineering. Project website: <<http://www.loa-cnr.it/DOLCE.html>>.
- 49 MASOLO, C.; BORGIO, S.; GANGEMI, A.; GUARINO, N.; OLTRAMARI, A. *Ontology Library: WonderWeb Deliverable D18*. Trento, Italy, 2003. Available at: <<http://www.loa-cnr.it/Papers/D18.pdf>>.
- 50 MAINTAINERS. General Formal Ontology. Project website: <<http://www.onto-med.de/ontologies/gfo/>>.
- 51 HELLER, B.; HERRE, H. Ontological categories in GOL. *Axiomathes*, v. 14, n. 1, p. 57–76, 2004.
- 52 MAINTAINERS. Basic Formal Ontology. Project website: <<http://ontology.buffalo.edu/bfo/>>.
- 53 GUIZZARDI, G.; WAGNER, G. Using the Unified Foundational Ontology (UFO) as a foundation for general conceptual modeling languages. Springer-Verlag, Berlin, 2009.
- 54 DEGEN, W.; HELLER, B.; HERRE, H.; SMITH, B. GOL: Toward an axiomatized upper-level ontology. In: *Proc. of the 2nd Int. Conf. on Formal Ontology in Information Systems*. New York, USA: ACM, 2001. p. 34–46. Ogunquit, USA. [doi: 10.1145/505168.505173].
- 55 LEVESQUE, H.; BRACHMAN, R. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, v. 3, n. 1, p. 78–93, 1987. [doi: 10.1111/j.1467-8640.1987.tb00176.x].
- 56 BAADER, F. et al. *The Description Logic handbook: Theory, implementation, and applications*. [S.l.]: Cambridge Univ. Press, 2003.
- 57 HORROCKS, I. et al. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics*, v. 1, n. 1, p. 7–26, 2003. [doi: 10.1016/j.websem.2003.07.001].
- 58 GUIZZARDI, G. The role of foundational ontologies for conceptual modeling and domain ontology representation. In: *Proc. of the 7th International Baltic Conf. on Databases and Information Systems*. [S.l.]: IEEE, 2006. p. 17–25. Vilnius, Lithuania. [doi: 10.1109/DBIS.2006.1678468].
- 59 CEUSTERS, W.; SMITH, B.; FLANAGAN, J. Ontology and medical terminology: Why Description Logics are not enough. In: *Proceedings of TEPR 2003 - Towards an Electronic Patient Record*. [S.l.: s.n.], 2003. San Antonio, USA.
- 60 GUIZZARDI, G.; GUARINO, N. An ontology-based approach for evaluating the domain appropriateness and comprehensibility appropriateness of modeling languages. In: *Proc. of the 8th International Conf. on Model Driven Engineering Languages and Systems (MoDELS)*. Berlin / Heidelberg: Springer, 2005. (LNCS, Volume 3713(2005)), p. 691–705. Montego Bay, Jamaica. [doi: 10.1007/11557432].

- 61 GUIZZARDI, G.; HALPIN, T. Ontological foundations for conceptual modeling. *Journal of Applied Ontology*, v. 3, n. 1-2, p. 1–12, 2008. [doi: 10.3233/AO-2008-0049].
- 62 USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, v. 11, n. 2, p. 93–136, 1996.
- 63 FALBO, R. A. Experiences in using a method for building domain ontologies. In: *Proc. of 16th Conf. On Software Engineering and Knowledge Engineering (SEKE'04)*. [S.l.: s.n.], 2004. p. 474–477. Banff, Canada.
- 64 GUARINO, N. Understanding, building and using ontologies. *Int. J. of Human-Computer Studies*, v. 46, n. 2-3, p. 293–310, 1997. [doi: 10.1006/ijhc.1996.0091].
- 65 JARRAR, M.; MEERSMAN, R. Ontology Engineering - The DOGMA approach. In: *Advances in Web Semantics I*. Berlin: Springer, 2008, (LNCS, v. 4891). p. 7–34. [doi: 10.1007/978-3-540-89784-2\_2].
- 66 van HEIJST, G.; SCHREIBER, A. T.; WIELINGA, B. J. Using explicit ontologies in KBS development. *International Journal of Human - Computer Studies*, v. 46, n. 2-3, p. 183–292, 1997. [doi: 10.1006/ijhc.1996.0090].
- 67 SMITH, B. et al. Relations in biomedical ontologies. *Genome Biology*, v. 6, n. 5, p. R46, 2005. [doi: 10.1186/gb-2005-6-5-r46].
- 68 MCCRAY, A. T. Conceptualizing the world: Lessons from history. *J. of Biomedical Informatics*, v. 39, n. 3, p. 267–273, 2006. [doi: 10.1016/j.jbi.2005.08.007].
- 69 CIMINO, J. J. In defense of the desiderata. *J. of Biomedical Informatics*, v. 39, n. 3, p. 299–306, 2006. [doi: 10.1016/j.jbi.2005.11.008].
- 70 SCHULZ, S.; KUMAR, A.; BITTNER, T. Biomedical ontologies: What part-of is and isn't. *J. of Biomedical Informatics*, v. 39, n. 3, p. 350–361, 2006. [doi: 10.1016/j.jbi.2005.11.003].
- 71 HOEHNDORF, R.; LOEBE, F.; POLI, R.; HERRE, H.; KELSO, J. GFO-Bio: A biological core ontology. *Applied Ontology*, v. 3, n. 4, p. 219–227, 2008. [doi: 10.3233/AO-2008-0055].
- 72 SCHULZ, S. et al. From GENIA to BIOTOP - Towards a top-level ontology for biology. In: BENNETT, B.; FELLBAUM, C. (Ed.). *Proc. of the 4th Int. Conf. of Formal Ontology in Information Systems (FOIS 2006)*. Amsterdam: IOS Press, 2006. (Frontiers in Artificial Intelligence and Applications, v. 150), p. 103–114.
- 73 RECTOR, A. Defaults, context, and knowledge: Alternatives for OWL-indexed knowledge bases. In: ALTMAN, R. B. et al. (Ed.). *Proc. of the 9th Pacific Symposium on Biocomputing (PSB 2004)*. Hawaii, USA: World Scientific, 2004. p. 226–237.
- 74 SCHULZ, S.; SUNTISRIVARAPORN, B.; BAADER, F.; BOEKER, M. SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International Journal of Medical Informatics*, 2008. [doi: 10.1016/j.ijmedinf.2008.06.004].
- 75 International Organization for Standardization. *ISO 1087-1: Terminology work - Vocabulary - Part 1: theory and applications*. Geneva, Switzerland, 2000.
- 76 BODENREIDER, O.; SMITH, B.; KUMAR, A.; BURGUN, A. Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies. *Artificial Intelligence in Medicine*, v. 39, n. 3, p. 183–195, 2007. [doi: 10.1016/j.artmed.2006.12.003].
- 77 MAINTAINERS. SNOMED-CT - Systematized Nomenclature of Medicine-Clinical Terms. Release January 2008. Project website: <<http://www.ihtsdo.org/snomed-ct/>>.
- 78 MCCRAY, A. T. An upper-level ontology for the biomedical domain. *Comparative and Functional Genomics*, v. 4, n. 1, p. 80–84, 2003. [doi: 10.1002/cfg.255].
- 79 SCULZE-KREMER, S.; SMITH, B.; KUMAR, A. Revising the UMLS Semantic Network. In: FIESCHI, M. et al. (Ed.). *Proceedings of the 11th World Congress on Medical Informatics*. San Francisco: IOS Press, 2004. (MEDINFO, Pt 1), p. 170–340.
- 80 KUMAR, A.; SMITH, B. The Unified Medical Language System and the Gene Ontology: Some critical reflections. Springer, Berlin / Heidelberg, Volume 2821/2003, p. 135–148, 2003. [doi: 10.1007/b13477].

- 81 GOLBECK, J.; FRAGOSO, G.; HARTEL, F.; HENDLER, J.; OBERTHALER, J.; PARSIA, B. The National Cancer Institute's thesaurus and ontology. *J. of Web Semantics*, v. 1, n. 1, p. 75–80, 2003.
- 82 SIOUTOS, N. et al. NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. *J. of Biomedical Informatics*, v. 40, n. 1, p. 30–43, 2007. [doi: 10.1016/j.jbi.2006.02.013].
- 83 CEUSTERS, W.; SMITH, B.; GOLDBERG, L. A terminological and ontological analysis of the NCI thesaurus. In: *Methods of Information in Medicine 2005*. [S.l.: s.n.], 2005. p. 498–507.
- 84 KUMAR, A.; SMITH, B. Artificial intelligence in medicine. In: . [S.l.]: Springer Berlin Heidelberg, 2005, (Lecture Notes in Computer Science, Volume 3581/2005). cap. Oncology ontology in the NCI thesaurus, p. 213–220. [doi: 10.1007/11527770].
- 85 GUARINO, N.; MUSEN, M. A. Applied Ontology: Focusing on content. *Applied Ontology*, v. 1, n. 1, p. 1–5, 2005.
- 86 CORNET, R.; KEIZER, N. de. Forty years of SNOMED: A literature review. *BMC Medical Informatics and Decision Making*, v. 8, n. Suppl 1, p. 1–6, 2008. [doi: 10.1186/1472-6947-8-S1-S2].
- 87 CIMINO, J. J. Terminology tools: State of the art and practical lessons. *Methods of Information in Medicine*, v. 40, n. 4, p. 298–306, 2001.
- 88 SMITH, B.; WILLIAMS, J.; SCHULZE-KREMER, S. The ontology of the Gene Ontology. In: *Proc. of the AMIA Symposium 2003*. [S.l.: s.n.], 2003. p. 609–13.
- 89 KUMAR, A.; SMITH, B.; NOVOTNY, D. Biomedical Informatics and granularity. *Comparative and Functional Genomics*, v. 5, n. 6-7, p. 501–08, 2004. [doi: 10.1002/cfg.429].
- 90 GUIZZARDI, G. The problem of transitivity of part-whole relations in Conceptual Modeling revisited. In: (Forthcoming) *Proc. of the 21st International Conf. on Advanced Information Systems Engineering (CAISE'09)*. [S.l.]: Springer, 2009. (LNCS). Amsterdam, The Netherlands.
- 91 MAINTAINERS. Gene Ontology Next Generation. Project website: <<http://www.gong.manchester.ac.uk/>>..
- 92 ARANGUREN, M. E.; WROE, C.; GOBLE, C.; STEVENS, R. In situ migration of handcrafted ontologies to reason-able forms. *Data & Knowledge Engineering*, v. 66, n. 1, p. 147–162, 2008. [doi: 10.1016/j.datak.2008.02.002].
- 93 LEWIS, S. E. Gene Ontology: Looking backwards and forwards. *Genome Biology*, v. 6, n. 1, p. 103.1–103.4, 2004. [doi: 10.1186/gb-2004-6-1-103].
- 94 MAINTAINERS. FMA - Foundational Model of Anatomy. Project website: <<http://sig.biostr.washington.edu/projects/fm/AboutFM.html>>..
- 95 RECTOR, A.; ROGERS, J.; BITTNER, T. Granularity, scale and collectivity: When size does and does not matter. *J. of Biomedical Informatics*, v. 39, n. 3, p. 333–349, 2006. [doi: 10.1016/j.jbi.2005.08.010].
- 96 DONNELLY, M.; BITTNER, T.; ROSSE, C. A formal theory for spatial representation and reasoning in biomedical ontologies. *Artificial Intelligence in Medicine*, v. 36, n. 1, p. 1–27, 2005. [doi: 10.1016/j.artmed.2005.07.004].
- 97 HAHN, U.; SCHULZ, S. Ontological foundations for biomedical sciences. *Artificial Intelligence in Medicine*, v. 39, n. 3, p. 179–182, 2007. [doi: 10.1016/j.artmed.2006.12.006].
- 98 BURGUN, A. A desiderata for domain reference ontologies in Biomedicine. *Journal of Biomedical Informatics*, v. 39, n. 3, p. 307–313, 2006. [doi: 10.1016/j.jbi.2005.09.002].
- 99 FERRARIO, R.; GUARINO, N. Towards an ontological foundation for Services Science. In: *Proc of the 1st Future Internet Symposium (FIS'08), Revised Selected Papers*. Berlin, Heidelberg: Springer-Verlag, 2009. p. 152–169. Vienna, Austria. [doi: 10.1007/978-3-642-00985-3\_13].
- 100 SCHULZ, S.; STENZHORN, H.; BOEKER, M.; KLAR, R.; SMITH, B. Clinical ontologies interfacing the real world. In: *3rd International Conference on Semantic Technologies (i-semantics 2007)*. Graz, Austria: [s.n.], 2007.

- 101 GENNARI, J. H.; SILBERFEIN, A.; WILEY, J. C. Integrating genomic knowledge sources through an anatomy ontology. In: *Proc of Pacific Symposium on Biocomputing*. [S.l.: s.n.], 2005. p. 115–126.
- 102 BLAKE, J. A.; RICHARDSON, J. E.; DAVISSON, M. T.; EPPIG, J. T. The Mouse Genome Database (MGD): A comprehensive public resource of genetic, phenotypic and genomic data. *Nucleic Acids Research*, Oxford University Press, v. 25, n. 1, p. 85–91, 1997. Available at: <<http://nar.oxfordjournals.org/cgi/content/short/25/1/85>>.
- 103 RUBIN, D. L.; DAMERON, O.; BASHIR, Y.; GROSSMAN, D.; DEV, P.; MUSEN, M. A. Using ontologies linked with geometric models to reason about penetrating injuries. *Artificial Intelligence in Medicine*, v. 37, n. 3, p. 167–176, 2006. [doi: 10.1016/j.artmed.2006.03.006].
- 104 DAMERON, O.; ROQUES, E.; RUBIN, D.; BURGUN, A. Grading lung tumors using OWL-DL based reasoning. In: *9th International Protégé Conference Proc.* [S.l.: s.n.], 2006.
- 105 RUBIN, D. L.; SHAH, N. H.; NOY, N. F. Biomedical ontologies: A functional perspective. *Briefings in Bioinformatics (Oxford)*, v. 9, n. 1, p. 75–90, 2007. [doi: 10.1093/bib/bbm059].
- 106 GUIZZARDI, G. et al. Grounding software domain ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology. In: *Proc. of the Iberoamerican Workshop on Requirements Engineering and Software Environments*. [S.l.: s.n.], 2008. p. 127–140. Recife, Brazil.
- 107 GUIZZARDI, G.; MASOLO, C.; BORGIO, S. In defense of a trope-based ontology for Conceptual Modeling: An example with the foundations of attributes, weak entities and datatypes. In: *Proc. of the 25th International Conf. on Conceptual Modeling (ER'06)*. Berlin / Heidelberg: Springer, 2006. (LNCS), p. 112–125. Tucson, USA. [doi: 10.1007/11901181].
- 108 GARDENFORS, P. *Conceptual spaces: the geometry of thought*. Cambridge, USA: MIT Press, 2000.
- 109 GUIZZARDI, G.; HERRE, H.; WAGNER, G. Towards ontological foundations for UML conceptual models. In: *Proc. of the Confederated International Conferences DOA, CoopIS and ODBASE*. [S.l.]: Springer, 2002. (LNCS, v. 2519), p. 1100–1117. Irvine, USA.
- 110 GUIZZARDI, G. Modal aspects of object types and part-whole relations and the *de re/de dicto* distinction. In: *Proc. of the 19th International Conf. on Advanced Information Systems Engineering (CAiSE'07)*. Berlin / Heidelberg: Springer, 2007. (LNCS), p. 5–20. Trondheim, Norway. [doi: 10.1007/978-3-540-72988-4\_2].
- 111 GUIZZARDI, G.; WAGNER, G.; GUARINO, N.; van SINDEREN, M. An ontologically well-founded profile for UML conceptual models. In: *Advanced Information Systems Engineering*. Berlin / Heidelberg: Springer, 2004. (LNCS, Volume 3084/2004), p. 112–126. [doi: 10.1007/b98058].
- 112 BENEVIDES, A. B.; GUIZZARDI, G. A model-based tool for conceptual modeling and domain ontology engineering in OntoUML. In: *(Forthcoming) Proc. of the 11th International Conf. on Enterprise Information Systems (ICEIS'09)*. [S.l.: s.n.], 2009. Milan, Italy.
- 113 MASOLO, C.; GUIZZARDI, G.; VIEU, L.; BOTTAZZI, E.; FERRARIO, R. Relational roles and qua-individuals. In: BOELLA, G. et al. (Ed.). *Roles, an Interdisciplinary Perspective: Ontologies, Programming Languages, and Multiagent Systems. Papers from the AAAI Fall Symposium*. Menlo Park, USA: AAAI Press, 2005. p. 103–112.
- 114 BUREK, P. *Ontology of Functions: A domain-independent framework for modeling functions*. PhD Thesis — University of Leipzig, Germany, 2006.
- 115 LOEBE, F. Abstract vs. social roles - towards a general theoretical account of roles. *Applied Ontology*, v. 2, n. 2, p. 127–158, 2007.
- 116 MAINTAINERS. OWL - Web Ontology Language. Project website: <<http://www.w3.org/TR/owl-features/>>.
- 117 MAINTAINERS. SWRL - Semantic Web Rule Language. Project website: <<http://www.w3.org/Submission/SWRL/>>.
- 118 HORROCKS, I. et al. OWL rules: A proposal and prototype implementation. *Journal of Web Semantics*, v. 3, n. 1, p. 23–40, 2005. [doi: 10.1016/j.websem.2005.05.003].

- 119 MOTIK, B. et al. Query answering for OWL-DL with rules. *Journal of Web Semantics*, v. 3, n. 1, p. 41–60, 2005. [doi: 10.1016/j.websem.2005.05.001].
- 120 PATEL-SCHNEIDER, P.; HORROCKS, I. A comparison of two modelling paradigms in the semantic web. *Journal of Web Semantics*, v. 5, n. 4, p. 240–50, 2007. [doi: 10.1016/j.websem.2007.09.004].
- 121 MAINTAINERS. Protégé OWL editor. Project website: <<http://protege.stanford.edu/>>.
- 122 ANTONIOU, G.; van HARMELEN, F. Web Ontology Language: OWL. In: STAAB, S.; STUDER, R. (Ed.). *Handbook on Ontologies*. [S.l.]: Springer, 2004, (Handbooks in Information Systems). cap. 4, p. 67–92.
- 123 WEINHAUS, A. J.; ROBERTS, K. Anatomy of the human heart. In: IAIZZO, P. (Ed.). *Handbook of cardiac anatomy, physiology, and devices*. Totowa, New Jersey: Humana Press, 2005. cap. 4.
- 124 SMITH, B. Fiat objects. *Topoi*, v. 20, n. 2, p. 131–148, 2001. [doi: 10.1023/A:1017948522031].
- 125 PRIBBENOW, S. Meronymic relationships: From classical mereology to complex part-whole relations. In: GREEN, R. et al. (Ed.). *The semantics of relationships: An interdisciplinary perspective*. [S.l.]: Springer, 2002, (Information Science and Knowledge Management, v. 3). cap. 3.
- 126 LASKE, T.; IAIZZO, P. The cardiac conduction system. In: IAIZZO, P. (Ed.). *Handbook of cardiac anatomy, physiology, and devices*. Totowa, New Jersey: Humana Press, 2005. cap. 9.
- 127 MAINTAINERS. openGALEN - Advanced terminology systems for clinical information systems. Project website: <<http://www.opengalen.org/>>.
- 128 GUYTON, A.; HALL, J. *Textbook of medical physiology*. 11th. ed. Philadelphia: Elsevier Saunders, 2006.
- 129 DUPRE, A.; VINCENT, S.; IAIZZO, P. A. Basic ECG theory, recordings, and interpretation. In: IAIZZO, P. (Ed.). *Handbook of cardiac anatomy, physiology, and devices*. Totowa, New Jersey: Humana Press, 2005. cap. 15.
- 130 GONCALVES, B.; ZAMBORLINI, V.; GUIZZARDI, G. An ontological analysis of the electrocardiogram. *Electronic Journal of Communication, Information and Innovation in Health. Supplement on Ontologies, Semantic Web and Health*, v. 3, n. 1, p. 907–28, 2009. Rio de Janeiro, Brazil. [doi: 10.3395/reciis.v3i1.242en].
- 131 HORRIDGE, M.; PATEL-SCHNEIDER, P. F. OWL 2 Web Ontology Language Manchester Syntax. Available at: <<http://www.w3.org/2007/OWL/wiki/ManchesterSyntax>>.
- 132 HORRIDGE, M. et al. The Manchester OWL Syntax. In: *Proc. of the 2nd OWL Experiences and Directions Workshop (OWLED'06)*. [S.l.: s.n.], 2006. Georgia, USA.
- 133 SIRIN, E. et al. Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics*, v. 5, n. 2, p. 51–53, 2007. [doi: 10.1016/j.websem.2007.03.004].
- 134 GONCALVES, B.; FILHO, J. G. P.; ANDREAIO, R. V.; GUIZZARDI, G. ECG data provisioning for telehomecare monitoring. In: *Proc. of the 2009 ACM symposium on Applied Computing (SAC'08)*. New York, USA: ACM, 2008. p. 1374–9. Fortaleza, Brazil. [doi: 10.1145/1363686.1364004].
- 135 GONCALVES, B.; FILHO, J. G. P.; ANDREAIO, R. V. EcgAware: An ECG markup language for ambulatory telemonitoring and decision making support. In: *Proc. of the International Conf. on Health Informatics*. [S.l.: s.n.], 2008. p. 37–43. Funchal, Portugal.
- 136 BASHSHUR, R. L.; MANDIL, S. H.; SHANNON, G. W. State-of-the-art Telemedicine/Telehealth: An international perspective. *Telemedicine Journal and e-Health*, v. 8, n. 1, 2002.
- 137 MAINTAINERS. PhysioNet. Project website: <<http://www.physionet.org/>>.
- 138 MOODY, G. B.; MARK, R. G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, v. 20, n. 3, p. 45–50, 2001.
- 139 CEN/TC-251. SCP Document CEN/TC-251 N02-15. Retrieved from: <<<http://www.centc251.org/>>>. August, 2006.
- 140 STOCKBRIDGE, N.; BROWN, B. Annotated ECG waveform data at FDA. *Journal of Electrocardiology*, v. 37, n. Supplement 1, p. 63–4, 2004. [doi: 10.1016/j.jelectrocard.2004.08.018].

- 141 BROWN, B.; BADILINI, F. HL7 aECG implementation guide. Available at: <[http://www.amps-llc.com/UsefulDocs/aECG\\_Implementation\\_Guide.pdf](http://www.amps-llc.com/UsefulDocs/aECG_Implementation_Guide.pdf)>. Access on March 15, 2009.
- 142 MOODY, G. B. WFDB programmer's guide. Version 10.4.19. Available at: <<http://www.physionet.org/physiotools/wpg/>>. Access on March 2009.
- 143 LAGUNA, P.; MARK, R. G.; GOLDBERGER, A.; MOODY, G. B. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. *Computers in Cardiology*, IEEE Computer Society Press, v. 24, p. 673–6, 1997.
- 144 MOODY, G. B.; MARK, R. G. Development and evaluation of a 2-lead ECG analysis program. *Computers in Cardiology*, IEEE Computer Society Press, p. 39–44, 1982.
- 145 MCBRIDE, B. Jena: A semantic web toolkit. *IEEE Internet Computing*, v. 6, n. 6, p. 55–59, 2002. [doi: 10.1109/MIC.2002.1067737].
- 146 ANDREA, R. V. et al. Ecg signal analysis through hidden Markov models. *IEEE Transactions on Biomedical Engineering*, v. 53, n. 8, p. 1541–9, 2006. [doi: 10.1109/TBME.2006.877103].
- 147 GOLDBERGER, A. et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, v. 101, n. 23, p. e215–e220, 2000.
- 148 WUNSCH, C. A. A toolkit for visualizing biomedical data sets. In: *Proc. of the 1st International Conf. on Computer graphics and interactive techniques in Australasia and South East Asia (GRAPHITE'03)*. New York, USA: ACM, 2003. p. 167–ff. Melbourne, Australia. [doi: 10.1145/604471.604505].
- 149 GARCÍA, R. et al. Interactive multimedia animation with macromedia flash in descriptive geometry teaching. *Computers & Education*, v. 49, n. 3, p. 615–639, 2007. [doi: 10.1016/j.compedu.2005.11.005].
- 150 GONCALVES, B.; ZAMBORLINI, V.; GUIZZARDI, G.; FILHO, J. G. P. An ontology-based application in heart electrophysiology: Representation, reasoning and visualization on the web. In: *Proc. of the 2009 ACM symposium on Applied Computing (SAC'09)*. New York, USA: ACM, 2009. p. 816–20. Hawaii, USA. [doi: 10.1145/1529282.1529456].
- 151 GONCALVES, B.; GUIZZARDI, G.; FILHO, J. G. P. An electrocardiogram (ECG) domain ontology. In: GUIZZARDI, G.; FARIAS, C. (Ed.). *Proc. of the 2nd Workshop on Ontologies and Metamodels for Software and Data Engineering (WOMSDE)*. [S.l.: s.n.], 2007. p. 68–81. João Pessoa, Brazil.
- 152 ZAMBORLINI, V.; GONCALVES, B.; GUIZZARDI, G. Codification and application of a well-founded heart-ECG ontology. In: *Proc. of the 3rd Workshop on Ontologies and Metamodels for Software and Data Engineering (WOMSDE)*. [S.l.: s.n.], 2008. Campinas, Brazil.
- 153 GONCALVES, B.; ZAMBORLINI, V.; GUIZZARDI, G.; FILHO, J. G. P. Using a lightweight ontology of heart electrophysiology in an interactive web application. In: *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web (WebMedia 2008)*. New York, USA: ACM, 2008. Vila Velha, Brazil.
- 154 ALLEN, J. F. Maintaining knowledge about temporal intervals. *Commun. ACM*, v. 26, n. 11, p. 832–843, 1983. [doi: 10.1145/182.358434].
- 155 PAN, F. *Representing complex temporal phenomena for the semantic web and natural language*. PhD Thesis — Computer Science Department, University of Southern California, 2007.
- 156 HOBBS, J. R.; PAN, F. Time Ontology in OWL. W3C Working Draft 27 September 2006. Project website: <<http://www.w3.org/TR/owl-time/>>.
- 157 BLOIS, M. S. Medicine and the nature of vertical reasoning. *The New England journal of medicine*, v. 318, n. 13, p. 847–51, 1988.
- 158 MCDERMOTT, D. Review to D. B. Lenat and R. V. Guha, Building large knowledge-based systems: Representation and inference in the CYC project. *Artificial Intelligence*, v. 61, n. 1, p. 53–63, 1993.
- 159 GUARINO, N. Formal Ontology, Conceptual Analysis and Knowledge Representation. *Int. Journal of Human and Computer Studies*, v. 45, n. 5-6, p. 625–40, 1995. [doi: 10.1006/ijhc.1995.1066].