Well Formed Clinical Practice Ontology Selection

David Mendes; Irene Pimenta Rodrigues

Departamento de Informática da Universidade de Évora

{dmendes;ipr}@uevora.pt

Abstract. We show how carefull shall an ontology selection process be in the specifig sub-domain of healthcare practice. This Ontology shall be well suited to reason about the clinical practice for it has to be based upon current Semantic Web techniques. Namely reasoners over OWL DL[1] ontologies have to handle the choosed ontology in such a way that the representational capabilities go hand-in-hand with adequate computability. We present the choice of ontology with all the theoretical considerations that have to be taken and show why the CPR[2] ontology is the best suited for our enrichment/population endeavours.

1 Introduction

We will present in this paper the reasons and causes of choosing CPR Ontology as the basis for our Clinical Practice domain Knowledge Representation.

1.1 Motivation

Since the early years of our century a large body of research has been developed in the Biomedical domain of knowedge. Beggining in 2006, the work around an ontology to adequately represent the healthcare providing activities has been around with a first proposal in 2009 as the CPR ontology. The Semantic Web tools and techniques have come of age to be able to use an ontology about a specific scientific and/or professional domain as knowledge representation scaffolding enough to be able to reason automatically and semantically inter-operate in that domain so we present here the selection process of such an ontology in a timely manner.

1.2 Previous work done

We are at the very beginning of the first author PhD work development under tutoring of the second. So far, only studying about the subject of Health Information knowledge representation and the Semantic Web tooling to reason around it has been done. To further develop and demonstrate the applicability of our work we have to choose and/or develop or enhance so we have been developing a carefull selection for a significant amount of time.

[1] Web Ontology Language - Description Logic

[2] Computer Based Record Ontology

2 What Ontology to Populate

Deciding what Ontology to Populate to function as the KB[3] to our Semantic Web Reasoning efforts is by itself a daunting task.

- The medical practice we want to represent is a many faceted science that renders a complex domain with issues to be addressed as
  - Temporality
  - Location
  - Granularity
  - High ambiguity in free text terminology
  - Jargon plagued with acronyms and even personal nicknames
- The Ontology shall take in consideration several different "best-practices" to be highly usable and used as intended
  - Solid design foundations for proper Ontology alignment and interoperability

  Well formed ontologies are able to support a variety of secondary uses not anticipated when the ontology was originally conceived [8]. In the process of trying to figure out the availability of such an ontology suitable for our purposes we found that the simplest was to develop an architectural software foundation to deliver them according to the Ontology Realism principles enunciated in [6] and with the freedom to be extendable according to anyone's particular needs. The ontologies here introduced that are to be in accordance to the OBO Foundry principles and thus interoperable may be a subset of any system brought up from our proposal. We just try to bring together the latest Software Engineering principles to the Ontology Engineering findings introduced in the referred article. With the loose coupling availability, configurable service inter-mixing, we picked what we could spot has low-hanging fruit to incorporate in our systems rendering them sub-optimal but demonstrable of the validity of the concepts and easily extendable/tunable with better ontology support and finer Web Service provisioning. For the moment the more widely accepted reference terminologies in form of ontologies that can be Web Service accessed through OntoCAT[4] or ODIE[5][6] were used and all the major coding standards that are similarly available were choosed. Given the impracticalities of using the whole UMLS, the relations and groups in the Semantic Network as preserved and only the MeSH[7][8] tractable terms and appropriate CORE views of SNOMED CT[9] are reached.

[3] Knowledge Base
[4] http://www.ontocat.org/
[5] Ontology Development & Information Extraction
[6] http://www.bioontology.org/ODIE
[7] Medical Subject Headings
[8] http://www.nlm.nih.gov/mesh/
[9] Systematized Nomenclature of Medicine - Clinical Terms

- Direction towards CSI[10]
Our coordinated, consistent ontologies shall be the structure for the Shared Meaning that is the most important concept to achieve CSI. The layer of understanding that is to be reached among disparate systems shall lie in an abstraction layer above the specific intricacies of each Clinical System. Considerations about the use of initiatives and already existing deliverables like HL7 V3 CDA[11], GreenCDA[12], CDISC[13] or BRIDG[14] are seriously considered.
- Integration for Extension
When trying to extend a particular given ontology to make it fit a particular purpose some techniques have been presented in the Ontology Engineering field. Some can be traced back to the early years of our century like Guarino and Welty [2]. Some new approaches are currently under heavy development and attracting special interest, these are mainly revolving around treating the ontologies as Metadata themselves and being able to process them for integration, clustering, validation or various other objectives. Work appearing recently is not reviewed here but we find projects like OASIS: Ontology Mapping and Integration Framework [9] and EL-VIRA [3] worth mentioning.
- Ontological Realism

Bayegan [1] proposed a process ontology for Clinical Practice which incorporated family-care workflow processes, clinical activities, different participants, and interactions of participants with a patient-record system. It was particularly interesting because it defined the minimal number of clinical headings necessary in a clinical setting, and also because their model was compatible with HL7. A similar effort was carried out by Scheuermann et al. [5], where they mainly focused on the disease and diagnosis manifestations. There are also some top-level ontologies in the literature (e.g., BFO[15] , BIOTOP, etc.) which can be further customized and expanded. Inspired by all these initiatives, W3C proposed an OWL-DL Computer-based Patient Record (CPR) ontology[16], which is briefly described in the following section 2.1

2.1 CPR

Computer-based Patient Record (CPR) was defined by the Institute of Medicine (IOM) in 1997 as an electronic patient record that resides in a system specifically designed to support users by providing accessibility to complete and accurate data, alerts, reminders, clinical decision support systems, links to medical knowledge, and other aids. Mostly they have been generally implemented and known

[10] Computer Semantic Interoperability
[11] Clinical Document Architecture
[12] http://wiki.hl7.org/index.php?title=GreenCDA_Project
[13] Clinical Data Interchange Standards Consortium
[14] Biomedical Research Integrated Domain Group
[15] http://www.ifomis.org/bfo
[16] http://code.google.com/p/cprontology

as Electronic Health Records (EHR) during recent years. The CPR Ontology addresses the terminology requirements of a CPR and its recording contents. These are defined as uniform core data elements, standardized coding systems and formats, a common data dictionary and information on outcomes of care and functional status. The ontology defines a minimal set of terms. It provides principled, ontological commitment for the terms used in many of the health-care information terminology systems. CPR relies on the use of foundational ontologies and ontology engineering best practices namely the OBO Foundry principles adherence are a requisite in its formation. Lastly it's intended to be used as an upper ontology of clinical medicine such as the OGMS[17]. In order to achieve uniformity, it needs to have significant coverage which turns into a pyramid ontology paradigm: small, well organized top and wide idiosyncratic bottom as seen in Figure 1

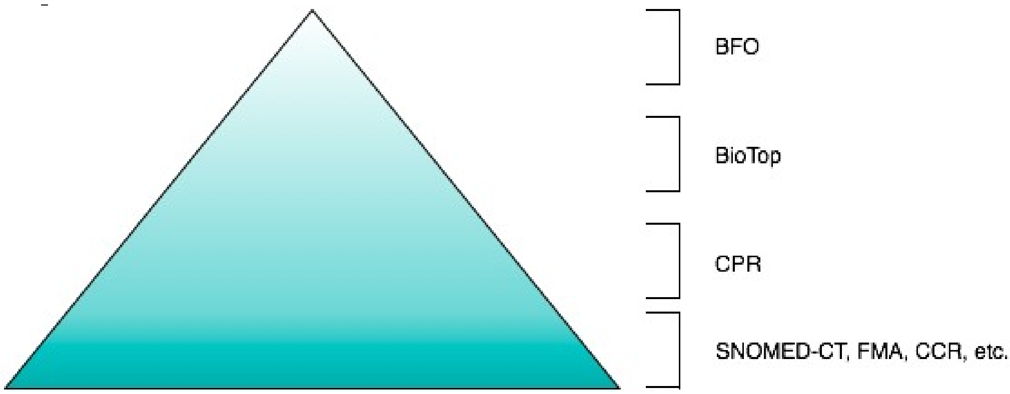

Fig. 1. CPR Pyramid

It adopts a cogent conceptual models that appeal to an ontological study of clinical medicine. High degree of care was taken into the development of CPR considering:

- The adoption of Ontological Realism has introduced by Barry et al. [7] according to the OBO Foundry principles
- Clear separation by definition of situations, findings and observables
- Differentiation among representational artifacts v.s. their referents
- Care act hierarchy and clinical workflow [1]
- Disease, diagnosis, etiology and the Disease Entity Model
- Disease, diagnosis, bodily features, etc. [5]
- Integrating anatomy, physiology, and pathology.

Use of realist ontology approach to the extent that distinctions are useful for real-world clinical informatics problems and validate against data and standard,

[17] Ontology for General Medical Science

controlled vocabularies namely SNOMED CT and FMA. There is a reasonable consensus around two reference ontologies that cover a substantial portion of clinical medicine: SNOMED-CT and the FMA. The location of equivalencies between classes and the extracted concepts from text is one of the major issues in our work so that the resulting populated ontology renders a realistic picture of the care process whose texts are the source of knowledge.

History and Motivation W3C first started to develop a Problem-Oriented Medical Record Ontology in 2006. The goal was to define a minimal set of health-care information terms while ontologically grounding HL7 RIM as a process model and using the criteria outlined in the traditional POMR structure W3C [11]. This led to the Web Ontology Language (OWL)-based ontology in November 2009, called the Computer-based Patient Record (CPR) ontology W3C [10]. Some parts of this ontology were taken from other top-level ontologies (e.g., BFO 1.1, BIOTOP, FMA, etc.) to ensure a sound and coherent means of necessary terminological representations required by an EHR. The surgical contributions to CPR has led into an ontology profoundly aligned with some basic "feeder ontologies" all of them according to OBO-Foundry principles and this alignment can de depicted as:

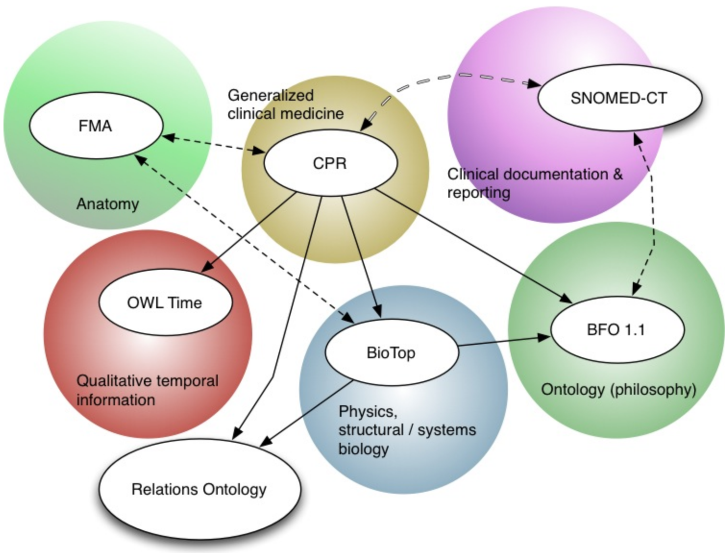

Fig. 2. CPR Ontology Alignment

Structure and Extensibility The main core concepts of this ontology are shown in the figure

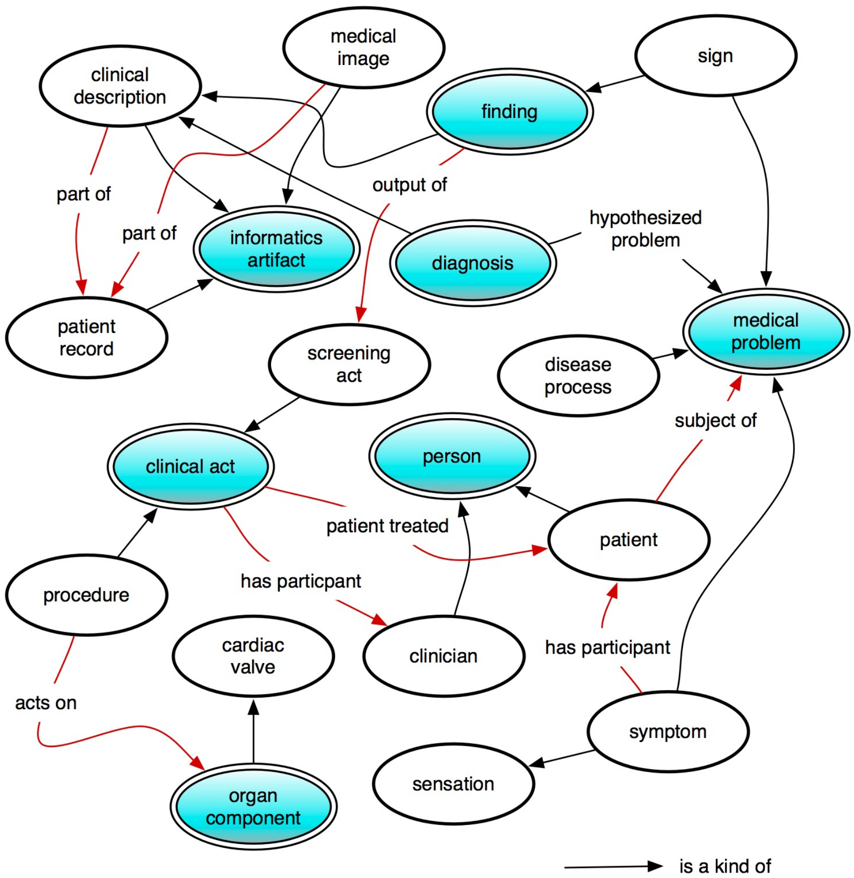

Fig. 3. Concepts of CPR ontology

2.2 Classes of CPR that are to be populated

With automated acquisition we extract information that will populate classes with creation of instances. As we shall see the information available is far from what should be considered like a minimum to render a full clinical practice ontology filling. Some trade-offs must be made and we shall squeeze all the texts in order to get as most as possible according to the CPR classes pre-defined. In the figure depicting CPR structure in 2.1 the top-level concepts of the CPR archetypes are shaded and shown with double circles. These are described below:

Clinical Acts: The most important concept of CPR ontology is Clinical Acts, which is used to model various clinical tasks and activities and the information flow in these activities. This ontology used the process ontology of defining clinical processes as a workflow model proposed by Bayegan et al. Bayegan [1] for defining the minimum clinical headings that are important for clinical communication and documentation. These clinical headings were put under the ‘span:Process’ class of BFO Ontology [45] to ensure proper classification of ocurrent and continuants data.There are four specializations of Clinical Acts: Clinical Administration Act, Clinical Investigation Act, Procedure, and Therapeutic Act. A Clinical Administration Act is defined as any administrative act which is not itself investigatory or therapeutic and is done for either the assessment or treatment (e.g., patient appointment). A Clinical Investigation Act is used to discover the status, causes and mechanisms of a patient’s health condition and is further classified into four classes: Clinical Analysis Act (used to generate the clinical hypothesis based on the condition of disease, physical examination, lab results, etc.), Diagnostic Procedure (the process of assessing the diagnosis; includes both laboratory or radiological procedures), Laboratory Tests (the process of quantitative or qualitative test of a substance in laboratory), Screening Act (collecting data from different aspects (e.g., clinical examination, medical history, social history, family history, etc.) to identify problems). A Procedure is a type of act which is taken to improve the patient’s condition. This concept is used in this ontology to incorporate both diagnostic and therapeutic procedures and is aligned with the definition of Procedure in HL7 RIM. Therapeutic acts are activities which are taken to improve or maintain the physical condition of a patient. This incorporates medical therapy (e.g., surgery), physical therapy (e.g., exercise), and psychological therapy (e.g., request to read an article that will improve the patient’s psychological status).

Medical Problems: In this ontology, medical problems are defined as entities which incorporate the signs, symptoms and confirmed diseases of a patient. Signs are abnormalities interpreted by clinicians during physical examinations whereas symptoms are particular sensations reported by the patient themselves. The disease process has been defined as either pathological disease or etiological agents while re-using the ontological framework for disease and diagnosis proposed by Scheuermann et al. Scheuermann et al. [5].

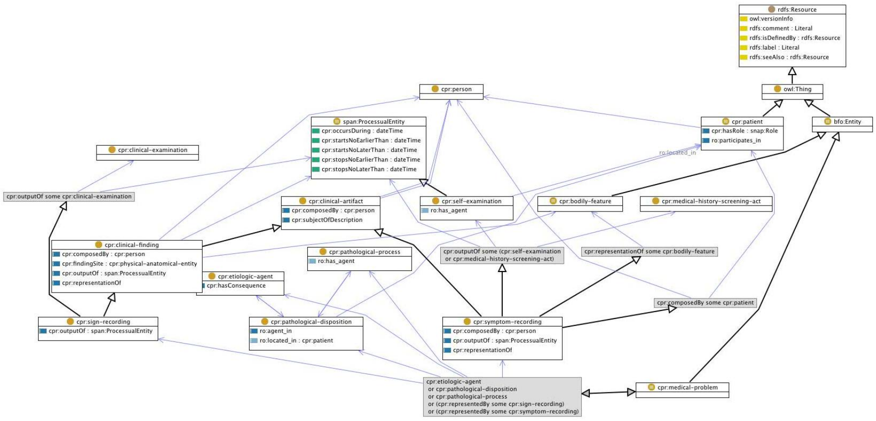

Fig. 4. CPR Medical Problems

This medical problems are, of course, an cpr:hipothesizedProblem of a cpr:clinical-diagnosis as we can see in the diagnosis view of the CPR ontology.

Findings: Findings are clinical examinations done by a clinical expert during an encounter to assess the condition of patient's body parts.

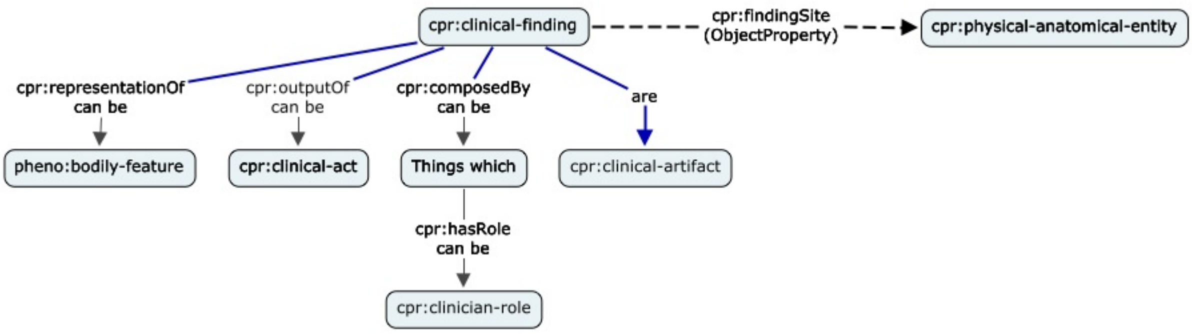

Fig. 5. Findings

Diagnosis: Diagnosis is not confirmed but hypothesized medical problem recorded during clinical analysis acts.

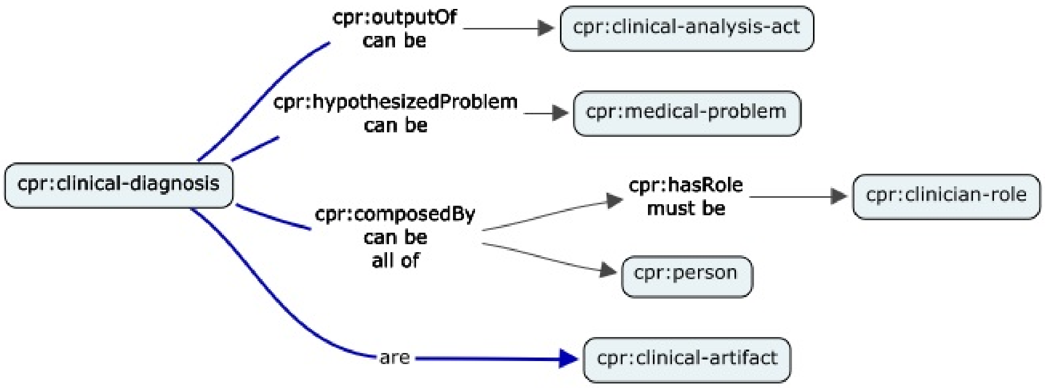

Fig. 6. CPR Diagnosis view

Informatics Artifacts: Informatics artifacts represent the pertinent information stored in an EHR. It includes all the clinical artifacts encountered in a patient, digital entities (e.g., diagnostic images), and other longitudinal information (e.g., clinical findings, symptoms). This concept is used to distinguish between the records of an action and the actual action itself.

Person: A person can be either the patient him- or herself or the clinically qualified person (e.g., nurse, general practitioner, etc.).

Organ Components: Organ components are the anatomical and pathological entities which take part in different clinical procedures and screening acts.

The CPR ontology is engineered in Protégé using OWL-DL language. Although it has all the necessary concepts an EHR should have, it lacks the properties of these concepts and the implementation of vocabulary binding in this ontology. To overcome this shortage we suggest the validation against the vocabularies that are translated partially to Portuguese and clinicians should be familiar with MeSHPOR.

2.3 CPR integrated with MeSHPOR

To define the properties of the concepts of this ontology a corresponding well defined and suitable vocabulary has to be adopted. Also, the vocabulary should be bound to this ontology so that the EHR concepts can use coded values where necessary like those that can be taken from the free text acquisition or from digging the EHR databases. We consider integrating for this purpose with the Portuguese localizations of MeSH[18] this version of the Medical Subject Headings is maintained and released annualy by the Latin-American and Caribbean Center on Health Sciences Information (Centro Latino-Americano e do Caribe

[18] http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MSHPOR

de Informação em Ciências da Saúde). The current version contains 26142 Main Headings from wich 14902 designated synonyms can be extracted. This wealthy resource of Portuguese translated terms will allow us, using a simple Ontological Engineering technique, to bind terms to created instances in CPR ontology.

3 Conclusion

We presented a humble contribution to demonstrate the cautions required to select a clinical practice Ontology suitable to be enriched/populated with instances collected automatically from reports taken from EHR or other colectable sources of information.

Bibliography

[1] Bayegan, E., 2002. Knowledge representation for relevance ranking of patient-record contents in primary-care situations. Ph.D. thesis, Norwegian University of Science and Technology, Faculty of Information Technology, Mathematics and Electrical Engineering, papers III and IV "This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.". 2, 2.1, 2.2

[2] Guarino, N., Welty, C., 2000. Identity, unity, and individuality: Towards a formal toolkit for ontological analysis. 2

[3] Hoehndorf, R., Dumontier, M., Oellrich, A., Wimalaratne, S., Rebholz-Schuhmann, D., Schofield, P., Gkoutos, G. V., 2011. A common layer of interoperability for biomedical ontologies based on owl el. Bioinformatics 27 (7), 1001–1008.
URL http://bioinformatics.oxfordjournals.org/content/27/7/1001.abstract 2

[4] Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., Musen, M. a., Jul. 2009. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research 37 (Web Server issue), W170–3.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2703982

[5] Scheuermann, R. H., Ceusters, W., Smith, B., 2009. Toward an Ontological Treatment of Disease and Diagnosis. In: 2009 AMIA Summit on Translational Bioinformatics. San Francisco, CA, pp. 116–120. 2, 2.1, 2.2

[6] Smith, B., Ceusters, W., Nov. 2010. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied ontology 5 (3-4), 139–188.
URL http://dx.doi.org/10.3233/AO-2010-0079 2

[7] Smith, B., Kumar, A., Ceusters, W., Rosse, C., Jan. 2005. On carcinomas and other pathological entities. Comparative and functional genomics 6 (7-8), 379–87.
URL http://www.pubmedcentral.nih.gov/ 2.1

[8] Smith, B., Scheuermann, R. H., Jan. 2011. Ontologies for clinical and translational research: Introduction. Journal of biomedical informatics 44, 3–7.
URL http://www.ncbi.nlm.nih.gov/pubmed/21241822 2

[9] Song, G., Qian, Y., Liu, Y., Zhang, K., 2006. Oasis: A mapping and integration framework for biomedical ontologies. In: Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems. IEEE Computer Society, Washington, DC, USA, pp. 611–616.
URL http://dl.acm.org/citation.cfm?id=1152999.1153011 2

[10] W3C, 2009. Compter-based patient record ontology.
URL http://code.google.com/p/cprontology 2.1

[11] W3C, 2009. Hcls pomrontology.
URL http://www.w3.org/wiki/HCLS/POMROntology 2.1