# UML INSIDE A KNOWLEDGE FRAMEWORK FOR ELECTRONIC PATIENT RECORDS

Anca Daniela IONIȚĂ[1], Cristina NICULESCU[2], Traian IONESCU[3]

*Articolul analizează complexitatea noțiunilor legate de dosarul electronic al pacientului şi diversitatea programelor dedicate, prezentând un cadru de cunoştinţe în care sunt incluse vocabulare medicale, terminologii, clasificări, metatezaure, reţele semantice, ontologii. Pe de o parte, aceste elemente sunt exemplificate prin standardele existente în domeniul medical; pe de altă parte, articolul propune un model de date sintetizator, reprezentat în UML (Unified Modeling Language), proiectat pentru un sistem de fişe electronice unificat. Acesta se leagă, prin anumite elemente, de terminologia şi clasificările standard din domeniul medical, făcând parte din cadrul de cunoştinţe prezentat în articol.*

*The paper analyzes the complexity of knowledge related to the electronic patient record and the diversity of dedicated software, presenting a knowledge framework containing medical vocabularies, terminologies, classifications, metathesauri, semantic networks and ontologies. On one side, these elements are exemplified with existent medical standards; on the other side, the paper proposes a synthesizing model, represented in UML (Unified Modeling Language), designed for a unified system of electronic records. This is connected, through some of its elements, to standard medical terminology and classifications, which are part of the knowledge framework presented in the paper.*

**Keywords**: Object Oriented Modeling, Knowledge Engineering, e-Health

## 1. Introduction

Electronic Patient Records (EPR) integrate information originated from various sources: hospitals, health centers, private clinics, general practitioners, laboratories, health insurance systems, screening programs and various centralized registers. The main purpose for keeping these records is to enhance the quality of the medical care, because one has a single access point to all items relevant for a certain patient [1]. It concerns medical and laboratory staff, insurers, and the patient himself

[1] Prof., The Department of Automatic Control and Industrial Informatics, University POLITEHNICA of Bucharest, Romania, e-mail: anca.ionita@aii.pub.ro

[2] Senior Researcher II, PhD, Research Institute for Artificial Intelligence, Romanian Academy, Bucharest, Romania, e-mail: ncristina@ racai.ro

[3] Prof., The Department of Automatic Control and Industrial Informatics, University POLITEHNICA of Bucharest, Romania e-mail: traiancionescu@ gmail.com

or herself. Thus, one eliminates the use of intermediate actors for obtaining paper charts with patient information, involved in direct discussions or telephone calls [2].

Scientific literature contains multiple concepts representing systems of patients' records, which have approximately the same meaning:

- **Personal Health Record (PHR)** - "records kept by a patient to make his or her navigation through the American healthcare system more efficient" [3];
- **Electronic Medical Record (EMR)** - "digital records kept by your doctor's office, your insurance company or the facilities where you are a patient" [3][2];
- **Computerized Patient Record (CPR) / Computer-based patient record (CPR)** - "an integration of patient information systems that captures and stores demographic, financial, and medical information from ancillary services such as registration, billing, lab, radiology, pathology, pharmacy, and transcription" [2];
- **Electronic Health Record (EHR)** - "a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports." [4];
- **Electronic Patient Record (EPR)** - "the record of the periodic care provided mainly by one institution" [5].

They all refer to the electronic collection of medical information, performed by the patient himself or herself, by a particular health-care institution, or by a global, integrated system. The current trend is to migrate towards the global, integrated option, for being able to deliver an improved care. For the purpose of this paper we shall refer to Electronic Patient Record as an integrating approach.

Electronic Patient Records have to collect and integrate a great quantity of non-homogenous data and information. Initially, information is collected as free text, like progress notes, which is very difficult to place in a certain context and to integrate with other pieces of information. A structured form serves better for the purpose of data integration. One can use, for example, a drop-down pick list, images, or digitized signals with associated metadata. However, structured data that use concepts or vocabularies not appropriate for the domain do not produce useful results.

EPRs are expected to improve quality of patient care [6], introducing rich details concerning each patient, access to up-to-date information, together with relevance of the collected data. Another advantage stands in more efficient tracking of patients and costs [7]; for instance, one can avoid repeating expensive tests [8].

An important challenge for integrating patient data from various heterogeneous sources is related to the multitude of standards in practical use, and the difficulty to support tool interoperation and data integration. The paper presents the diversity of existent EPR systems (chapter 2), and of standard medical vocabulary, terminology and classifications (chapter 3), then it introduces a framework for analyzing medical knowledge representations (chapter 4). Finally, one describes a UML representation of EPR data model (chapter 5), designed as a unifying conceptual model that can be subsequently linked to various standards in use. The

mappings between one of the UML class diagrams and various standard knowledge organizers are indicated in chapter 6.

## 2. The diversity of EPR systems

Early EPR systems projects were developed by academic medical centers; they first introduced the ideas and technologies that are currently used in Commercial-off-the shelf (COTS) systems, but they included non-standard vocabularies and interfaces. Some important early projects are presented in [9] and are summarized in Table 1.

*Table 1*

**Early EPR systems**

| Acronym | Name | Institution | Information |
|---|---|---|---|
| CHCS | Composite Health Care System | Department of Defense's (DoD) | - used worldwide |
| COSTAR | Computer Stored Ambulatory Record | Harvard | - placed in the public domain in 1975 |
| DHCP | De-Centralized Hospital Computer Program | Veteran's Administration | - used in all USA |
| HELP | Health Evaluation through Logical Processing | Latter-Day Saints Hospital, University of Utah | - pioneering decision support features |
| TDS | | Lockheed | |
| THERESA | | Grady Memorial Hospital, Emory University | - encouraging direct physician data entry |
| TMR | The Medical Record | Duke University Medical Center | |

Many companies provide electronic systems for patient records, sometimes called by their interchangeable names explained in the introduction. They are designed to assist providers in improving patient outcomes, and reducing medical errors and costs, but generally they are used in particular healthcare institutions and not at a global scale. Moreover, there are significant financial barriers in the way of adopting the technology: the necessity of a large capital investment; the difficulty to earn government incentive payments, lack of IT facilities in some rural areas. Below,

Table 2 presents a list with some proprietary EPR systems available on the market.

*Table 2*

**List of proprietary systems for electronic patient recording**

| Product | Company | Features |
|---|---|---|
| Centricity EMR | DataFuzion | - access for ambulatory care physicians and clinical staff to documenting patient encounters and clinical workflows<br>- secure clinical data exchange |
| Centricity™ | GE | - collaboration across the entire continuum of care |

| Enterprise | Healthcare | |
|---|---|---|
| Centricity™ Clinical Information Systems | GE Healthcare | - patient information on-line and at the point of care.<br>- planning care and electronic patient records.<br>- components for critical care, anesthesia and operating room management |
| Soarian Integrated Care | Siemens | - a web-based eHealth solution for the communication of patient-related data, forms and documents |
| Practice Partner® Patient Records | McKesson | - feature set utilized nationwide by medical offices<br>- prescription writing, faxing and drug interaction/allergy checking |
| Horizon Ambulatory Care™ | McKesson | - workflow-driven ambulatory electronic health record (EHR) solution that automates the clinical functions<br>- shared patient information |
| Integrated Medical Records (IMR) | IBM | - middleware used to integrate and correlate medical records from diverse sources and transform data into knowledge |
| Mobile Clinical Assistant (MCA) | Intel | - access up-to-date patient care records at the point of care and to enable documentation of a patient's condition in real time |
| HealthVault | Microsoft | - store and track personal health information online |

Moreover, Table 3 presents a list of software packages and applications licensed under an open source license or in the public domain for use in the healthcare industry, for electronic health or medical records.

*Table 3*

**List of open source electronic patient records**

| Licence | Name | Description |
|---|---|---|
| GNU Public License (GPL) | FreeMED | - tracking medical data, in detail, with preservation not just of the diagnosis but the reasons for medical encounters |
| | GNUmed | - a free software/open source medical practice management<br>- respects the privacy of patients |
| | Hospital OS | - managing hospital operations<br>- a Client - Server software |
| | HOSxP | - hospital information system, including EHR<br>- aims to ease the healthcare workflow |
| | OpenEMR | an open-source Electronic Medical Record (EMR) software |
| | OSCAR McMaster | - the billing component specialized for Canada health care providers. |
| Mozilla Public License (MPL) | Mirth | - bi-directional sending of HL7 messages |
| | OpenMRS | - extensible and scalable EMR based on Java |
| | THIRRA (EHR) | - a web based EHR application<br>- includes communicable diseases biosurveillance feature. |
| Apache Software License | ZEPRS | - an electronic patient record system that enables clinicians to enter data from patient visits using a web browser. |

## 3. The diversity of standard medical vocabulary, terminology, classifications

EPRs use both technical and clinical standards. The use of standard clinical vocabularies and ontologies (with structured data organization and relations to each other) significantly enhances the ability of clinical systems to interoperate, and allows patient related data to be used in clinical trials [9]. The main organizations entitled to adopt health systems standards are: *Health Level Seven* (HL7), *European Committee for Standardization* (CEN - Comité Européen de Normalisation), *American Society for Testing and Materials* (ASTM) E31. According to [10] some other standardization organizations oriented on health issues are: *Consolidated Health Informatics Initiative* (CHI) and *Clinical Data Interchange Standards Consortium* (CDISC).

An important task of standardization organizations is medical classification or coding, i.e. the assignment of numeric codes to each diagnosis, procedure and service, in order to store information in multiple repositories and then be able to retrieve it easily for the purpose of statistics, service payment, research, planning and foresight. Standards are created starting from key medical terms, which are identified and abstracted from patient records and they are equally useful for hospitals, physician's offices, home health care, long term care and insurance firms. Some important resources for medical classification are described below.

**Current Procedure Terminology (CPT) Codes** represents a medical nomenclature for health procedures and services. CPT Codes are maintained by American Medical Association (AMA) and are updated three times a year. AMA also offers products and services for practical use of CTP codes, like: CPT Network (a database for accurately billing and coding reimbursements) and CPT/RVU Search (a tool for getting details about the relative value payment amount associated with CTP codes).

**International Classification of Diseases (ICD)** is also used to code diseases or problems for billing purposes. It is adopted by World Health Organization (WHO), as the international standard of epidemiological, health management purposes and clinical diagnostic classification. It is based on the study of the general health situation for population groups, analyzing the incidence of various health problems in respect with individual circumstances, reimbursement, quality and guidelines [11].

**Diagnosis Related Groups (DRG)** is a system of classifying and distributing patients into approximately 500 groups, on the basis of their diagnosis treatment procedures, and other general criteria like age or gender [12]. Moreover, DRGs may be further grouped into 25 Major Diagnostic Categories (MDCs),

which are diagnosis areas, associated with a single medical specialty or with a single organ system or etiology.

**Logical Observation Identifiers, Names and Codes (LOINC)** is a universal standard used for the medical terminology of EPRs, including observations regarding laboratory results, clinical observations, and diagnostic studies. It contains specialties like chemistry, toxicology, hematology, microbiology; surgical pathology etc. [13].

**Systematized Nomenclature of Medicine (SNOMED)** is defined for indexing the entire EPR. SNOMED - CT (Clinical Terms) [14] includes more than 311,000 hierarchically organized concepts, with approximately 1,360,000 links or semantic relationships between them.

**Unified Medical Language System (UMLS)** [15],[14] created by U.S. National Library of Medicine (NLM), is based on a knowledge base which can be used for supporting EPRs. It is based on three components: a **Metathesaurus** (a vocabulary database available in multiple languages, covering concepts related to biomedicine and health, together with the relationships among them), a **Semantic Network** (providing a classification basis for Metathesaurus concepts, formed of more than a hundred types), a **SPECIALIST Lexicon** (including lexical information characterized from three points of view: syntactic, morphological, and orthographic).

Besides diagnosis and treatment procedures, reconciliation between various health centers is also based on the use of a standardized nomenclature of medication [16]. This is also important for e-prescriptions and for interoperation with pharmacy systems. Thus, **RxNorm** represents a nomenclature for clinical drugs adopted by National Library of Medicine (NLM) and used in US federal government systems for e-health [17]. Another international standard, especially used in Europe, is the **Anatomical Therapeutic Chemical (ATC)** classification system and the **Defined Daily Dose (DDD)** [18].

## 4. Framework for Analyzing Medical Knowledge Representations

Fig. 1 presents a knowledge framework, which was used to analyze medical knowledge necessary to develop an *Electronic Patient Record*. The vocabularies, classifications and terminologies presented in chapter 3 are grouped as knowledge organizers. Supplementary, there are two groups which are important for supporting medical software, discussed below: semantic knowledge (with medical domain ontologies and semantic networks) and modeling languages (with general and domain specific languages). The three groups influence one another and implementations must integrate them for obtaining a complete view of the medical system.

An ontology can be viewed as a declarative model of a domain that defines and represents the concepts existing in that domain, their attributes and the relationships between them. It is typically represented as a knowledge base which then becomes available to applications that need to use and/or share the knowledge of a domain. Within health informatics, an ontology is a formal description of a health-related domain [20].

The use of ontologies in medicine is mainly focused on the representation and (re-)organization of medical terminologies. Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Medical information systems, on the other hand, need to be able to communicate complex and detailed medical concepts (possibly expressed in different languages) unambiguously. This is obviously a difficult task and requires a profound analysis of the structure and the concepts of medical terminologies. But it can be achieved by constructing medical domain ontologies for representing medical terminology systems; some important examples are: the Foundational Model of Anatomy, the Galen-Core high-level ontology, MedO biomedical ontology, the ontology for HL7 Reference Information Model (RIM) [20]. Ontology-based applications have also been built in the field of Medical Natural Language Processing [20].

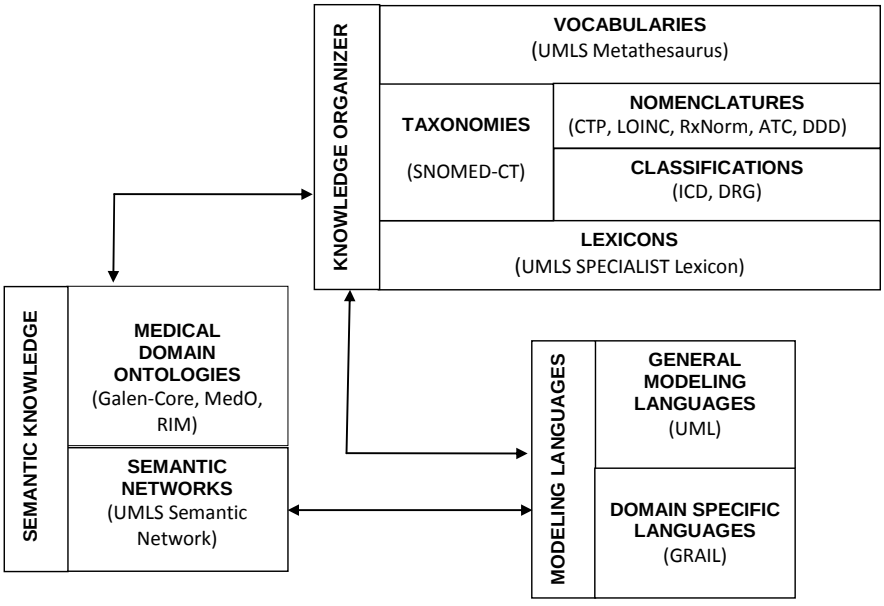


Fig. 1 Medical Knowledge Framework

Moreover, the development of medical applications is based on models expressed in various languages, which grasp the logic view (containing the most important concepts and the relationships between them) but hide the huge amount of details characteristic to the domain. For the conformity with existent knowledge (i.e. nomenclatures, classifications, vocabularies, taxonomies, lexicons) there should be modeling elements that establish bridges with knowledge organizers and foster the use of such standards in e-Health applications. The framework presented in **Fig. 1** includes general modeling languages, like UML – subsequently used in this paper, and also domain specific languages (DSLs), like GRAIL. The latter is a concept modeling language for developing computer-based medical terminologies. It is a compositional and generative semantic network knowledge representation language. The GRAIL representation scheme forms the basis of the GALEN Common Reference Model [21].

## 5. Proposed Data Model for Electronic Patient Record

Based on the multiple knowledge sources currently available, our study produced a data model for EPR, containing 48 classes, grouped in 8 packages, each one represented through class diagrams that contain detailed information, exposed as attributes, operations, generalizations and other kinds of relationships. This model respects the specifications of direct care functions for electronic health records, adopted as a HL7 (Health Level Seven) standard [19].

The Electronic Patient Record System modeled here was designed both for healthcare providers and patients. They should be able to retrieve important information – essential for delivering a high quality, integrated care [19]. Therefore, for each patient there should be a single logical health record, managing his or her demographics, data generated from various sources (including patient originated), health history, and progressing through consents, assessments, care plans, orders, and results.

The EPR data model is organized in 7 packages (see **Error! Reference source not found.**) - each one described by class diagrams that illustrate the most important classes, their attributes, operations and relationships. The ERP packages are:

- **Record Management package** – represents EPR core package and models structured and unstructured data related to the patient; it merges the following sub-packages:
- **Demographic Information package –** includes clinically relevant data, characterizing the patient as a member of a human population segment;

- **Health Characterization package** – includes a description of problems and medications delivered to the patient, together with the background of allergic and adverse reactions;
- **Permissions package** – represents preferences, directives, consents and authorizations captured from patients and their families;
- **External Data package** – includes image documents and other clinically relevant data from external clinical sources;
- **Medical Care package** – presents information related to various episodes of care, with their associated documentation, and also summarizes all prescriptions, medication orders, orders to transfer patients between healthcare units, diagnostic tests and laboratory orders.
- **Order Management package** – contains medical orders sent to other healthcare professionals for the patient's treatment;
- **Document Management package** - includes classes related to the documentation of patient care, like test results, clinical measurements and various clinical documents.

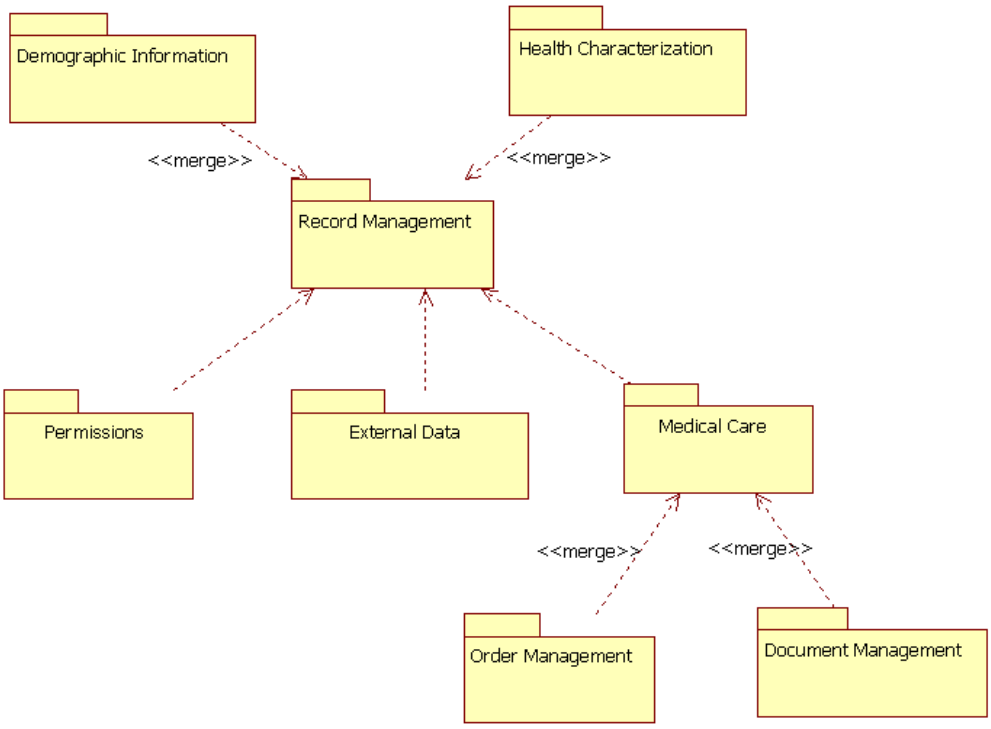


Fig. 2 Packages of Electronic Patient Record System

## 6. Mapping the UML Model on Standard Knowledge Organizers

For the purpose of our paper, we describe the Health Characterization package, presented in *Fig. 3*, and some mappings between its concepts and some standard medical vocabulary, terminology and classifications presented above. It contains classes for the patient characterization from the medical point of view, including information about the allergic background, the history of health problems suffered in the course of time, and the medication received for solving these problems.

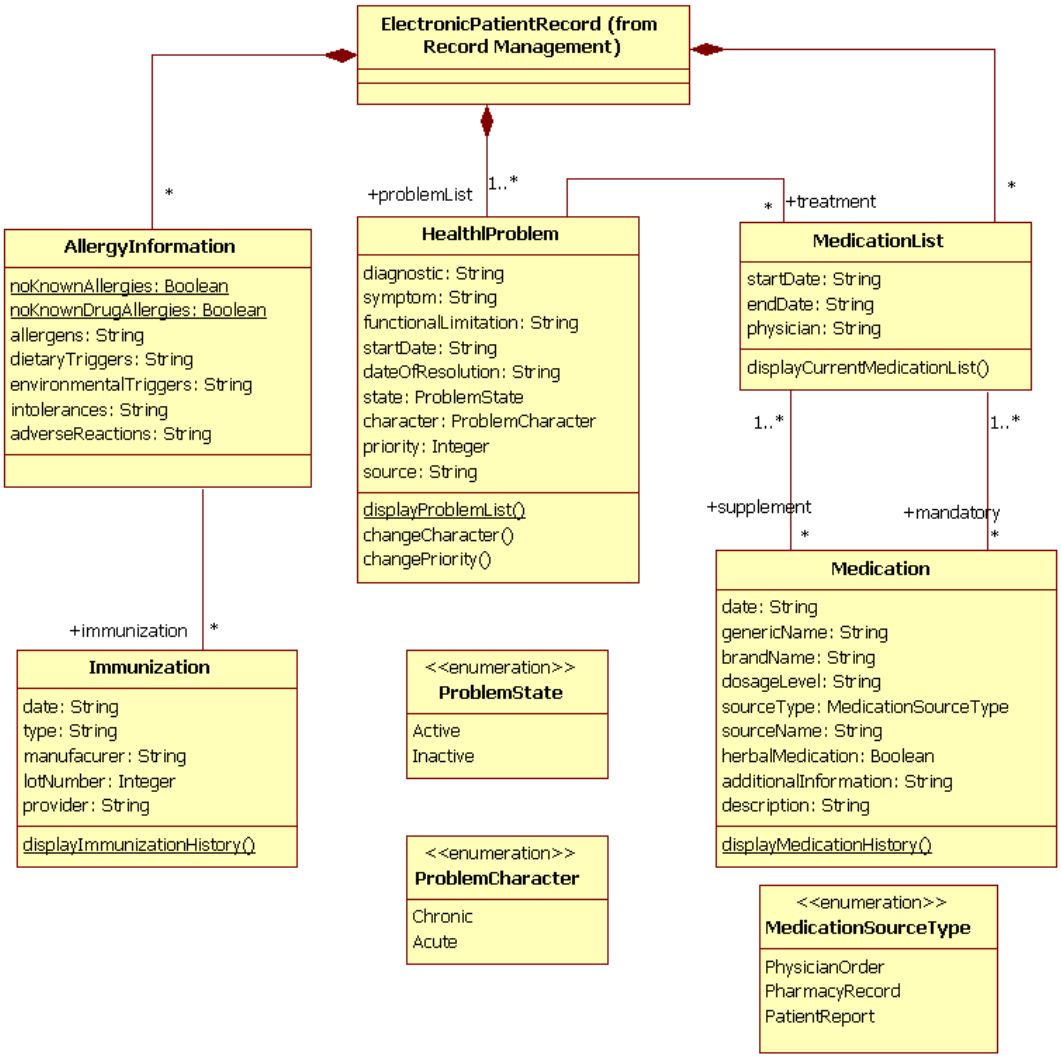


Fig. 3 Medical Characterization Package

*HealthProblem* represents the most important class of this package and it is essential for EPRs, because complete descriptions of diagnosis are recorded over time, characterized by the list of symptoms or functional limitations, the chronic or acute nature, the active / inactive current state. The link with various standards from our framework is made through the diagnosis specification, which can be associated with an IDC code, a DRG category, or an element of UMLS Metathesaurus.

The *AllergyInformation* class specifies any previously reported allergies, intolerances and adverse reactions, accumulating the entire patient's history; it makes the difference between various triggers, related to drugs, diet or environment. The *Immunization* class is used for maintaining a list of active immunizations or vaccination information, including the type of diseases the patient is immunized against.

The *MedicationList* class contains a list of medications associated to a certain clinical episode or health problem, each of them being an instance of *Medication* class, which characterizes any licensed drug used for treating or curing a patient's disease, health problem, symptoms or functional limitations. It contains the generic name, the brand name, dosage levels, and additional information regarding adverse effects, side effects or contraindications. This information corresponds to the up-to-date knowledge regarding the medicine, generally stored in RxNorm, or ATC/DDD nomenclatures.

## 7. Conclusions

The paper analyzed diverse knowledge representations that help one for keeping a centralized Electronic Patient Record or for integrating various distributed record systems. EPR software has to face many challenges of conformance and interoperation, due to the existence of multiple standards and classifications, and the complexity of the domain. The UML data model presented here has the advantage to hide the details at first glace, and to allow their further introduction, through mappings on other standardized representations of the medical domain knowledge.

## REFERENCES

[1] *J. Powell*, Data sources and reporting, in O.M. Jensen, D.M. Parkin, R. MacLennan, C.S. Muir and R.G. Skeet Eds., Cancer registration: Principles and Methods, IACR Scientific Publications No 95, 1991, pp. 29-42

[2] *T.L. Erstad*, Analyzing Computer based Patient Records: A Review of Literature, Journal of Healthcare Information Management, **Vol. 17**, No. 4, pp. 51-57

[3] *T. Torrey*, What is an EMR (Electronic Medical Record) / EHR (Electronic Health Record)? One Type of Digital Patient Health and Medical Record, February 07, 2008, About.com Guide

[4] *Jennifer Allan, and Jane Englebright*, "Patient-Centered Documentation: An Effective and Efficient Use of Clinical Information Systems," *Journal of Nursing Administration*. 30(2):90-95, February 2000
[5] *R. Beaumont*, Managing your Clinical Information, The Electronic Patient and Healthcare Record
[6] *Clifford Goodman*, Perspective; Savings in Electronic Medical Record Systems? Do It for the Quality. It is Unrealistic to Hold Out Widespread Adoption of Health Information. The People-To-People Health Foundation, Inc., September 2005 – October 2005.
[7] "How Technology is Transforming Your Hospital – Finally." *Business Week*. March 28, 2005
[8] *Gary Baldwin*, "Big IT Spenders." HealthLeaders News, November 11, 2004
[9] MITRE Corporation, Electronic Health Records Overview, April 2006, Available at: www.MITRE.org
[10] *A.S. Apostol, C. Câtu, V.C. Vernic*, Electronical Health Record's Systems. Interoperability, Annals. Computer Science Series. 1st Tome 1st Fasc., pp. 7-20, 2009
[11] World Health Organization, WHO Business Plan for Classifications … building blocks of health information , Version 1.0, Geneva 2005
[12] List Of Diagnosis Related Groups (DRGS), FY 2005, http://www.cms.gov/MedicareFeeforSvcPartsAB/Downloads/DRGDesc05.pdf
[13] *C. McDonald, S. Huff, K. Mercer, A. Hernandez, D.J. Vreeman*, Logical Observation Identifiers Names and Codes (LOINC®) Users' Guide, December 2010, http://loinc.org
[14] International Health Terminology Standards Development Organisation, SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms), http://www.ihtsdo.org/snomed-ct/
[15] Unified Medical Language System® (UMLS®), http://www.nlm.nih.gov/research/umls/about_umls.html
[16] *S. Liu, Ma Wei, R. Moore, V. Ganesan, S. Nelson*, RxNorm: prescription for electronic drug information exchange, IT Professional, **Vol. 7**, Issue:5, pp. 17 – 23, 2005
[17] *J. Kapusnik-Uner*, Medication Interoperability and a Means to Achieving Meaningful Use: RxNORM Developers and Early Adopters, Proc. Of AMIA 2010 Symposium, pp. 1369-1370, 2010
[18] WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health, Guidelines for ATC classification and DDD assignment, 14th Edition, 2011
[19] HL7 EHR TC, Electronic Health Record – System Functional Model, Chapter Three: Direct Care Functions, Release 1, February 2007
[20] Open Clinical: Ontologies, http://www.openclinical.org/ontologies.html
[21] Open Clinical: Software downloads: medical terminologies, http://www.openclinical.org/dld_galenGrail.html.