Multi-layer Data Model

Doctoral Thesis

Michal Huptych

Prague, January, 2013

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Artificial Intelligence and Biocybernetics

Supervisor: Doc. Ing. Lenka Lhotská, CSc
Acknowledgement

I would like to thank to my supervisor professor Lenka Lhotská for all of the great support, help and patience during the years. My thanks belong also to all my colleagues in the Biomedical Data and Signal Processing Group for inspiring work environment. Further, I would especially like to thank to Mikuláš Mlček for the opportunity to participate at the great work and his endless enthusiasm.

Last but not least, I would like to express my gratitude and thanks to my family for their support and understanding during my studies.

This work was supported by the research program No.NT11124-6/2010 “Cardiotocography evaluation by means of artificial intelligence” of the Ministry of Health Care and the research program "Information Society" under grant No. 1ET201210527 "Knowledge-based support of diagnostics and prediction in cardiology".
Abstract

With rapid development of IT and mobile technologies and their use in the medical field, it is necessary to guarantee the communication and representation of data in a much clearer form than ever before. Today, there are a number of standards that define syntactic form and structure of communication and representation as well as a number of classifications and dictionaries in a hierarchical arrangement, or even as an ontology, defining the content of a wide range of data communication and information exchange. In many aspects, these approaches allow semantic representation of data, however in the domain of signals and their integration with other data such representation is not yet resolved. In this thesis we present a proposed solution to representation of the signal on a higher level with the definitions of signal integration with other data from the medical record. We have created an ontology and architecture of multi-layer data model that allows the definition of complex signal, with indication of characteristic (important) points and complexes and with link to clinical events and background knowledge base. The proposed model has been applied to and verified in clinical practice in obstetrics, and in medical experimental area.
Abstrakt

Se strmým rozvojem IT a mobilních technologií a jejich využívání v oblasti zdravotnictví, je zapotřebí zaručit komunikaci a reprezentaci dat v mnohem jasnější formě než kdy předtím. Dnes již sice existuje řada standardů, které definují jak syntaktickou formu a strukturu komunikace a reprezentace tak také řada číselníků a slovníků, v hierarchickém uspořádání nebo dokonce jako ontologie, pro definici celé řady údajů obsahu komunikace a výměny informací. V mnoha ohledech umožňují tyto přístupy plnou reprezentaci, avšak v oblasti signálů a jejich propojení s ostatními údaji tato reprezentace dořešena není. V této práci předkládáme návrh řešení pro reprezentaci signálu na vyšší úrovni s definicemi jejich propojení s ostatními údaji ze zdravotnického záznamu. Vytvořili jsme architekturu vícevrstvého datového modelu, který umožňuje komplexní definici signálu, s označením charakteristických a důležitých bodů a komplexů s propojením na bázi doménových znalostí a propojený s událostmi. Navržený model je prezentován v oblasti klinické praxe, v oblasti porodnicvtví, a v lékařské experimentální oblasti.
Contents

1. INTRODUCTION ................................................................. 1
   1.1. Thesis contributions .................................................... 2
   1.2. Thesis aims .............................................................. 3
   1.3. Thesis outline ........................................................... 4

2. KNOWLEDGE REPRESENTATION AND ONTOLOGY ............ 7
   2.1. Data, Information and Knowledge .................................. 8
   2.2. Knowledge Representations ......................................... 10
      2.2.1. Propositional Logic .............................................. 10
      2.2.2. Predicate Logic ................................................ 11
      2.2.3. Description Logic .............................................. 11
      2.2.4. Rules ............................................................. 12
      2.2.5. Semantic Networks ............................................ 13
      2.2.6. Frames ............................................................ 13
   2.3. Ontology ................................................................... 14

3. INTEROPERABILITY, STANDARDS AND INFORMATION
   SYSTEMS IN MEDICINE ....................................................... 19
   3.1. Interoperability ......................................................... 19
      3.1.1. Technical / syntactic interoperability ....................... 20
      3.1.2. Semantic interoperability ..................................... 20
      3.1.3. Process interoperability ....................................... 21
   3.2. Standards and nomenclatures in medicine ..................... 22
      3.2.1. Structural Standardization ..................................... 22
      3.2.2. Content Standardization ...................................... 23
      3.2.3. HL7 standard ..................................................... 24
      3.2.4. CEN TC 251 a ENV 13606 .................................... 27
      3.2.5. DICOM ............................................................ 28
      3.2.6. ICD 10 ............................................................ 29
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2</td>
<td>Assumptions of the system</td>
<td>80</td>
</tr>
<tr>
<td>7.3</td>
<td>Design and Implementation of Applications</td>
<td>81</td>
</tr>
<tr>
<td>7.4</td>
<td>Events</td>
<td>83</td>
</tr>
<tr>
<td>7.6</td>
<td>Class Diagram of Multi-layer Model</td>
<td>93</td>
</tr>
<tr>
<td>8</td>
<td>MAIN CONTRIBUTION AND ACHIEVEMENTS</td>
<td>95</td>
</tr>
<tr>
<td>9</td>
<td>CONCLUSION</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY</td>
<td>103</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1 Function of Knowledge and Data in Enterprises [1] .............. 8
Figure 2 Connection of data, information and knowledge by [2] .......... 9
Figure 3 Example of a rule from MYCIN expert system [3] .............. 12
Figure 4 Semantic network example ....................................... 13
Figure 5 Example of frame ..................................................... 14
Figure 6 Simple example of Ontology - ontology of wine [8] .......... 16
Figure 7 HL7 RIM core classes, core attributes and relationship between classes [23] ............................................................... 25
Figure 8 a) Point to Point HL7 interface, b) HL7 Messages Server [27] .... 26
Figure 9 EHCR Development Levels according to Medical Record Institute [23] .......................................................... 32
Figure 10 Example of HL7 V3 XML message ................................ 36
Figure 11 Ontology of Labour .................................................. 44
Figure 12 Use case diagram of process admission to maternity hospital, labour and admission to neonatology .................................... 46
Figure 13 Use case diagram of admission to maternity hospital ......... 48
Figure 14 Use case diagram of Labour process with relation to observation, diagnosis and intervention (medication and surgery) .......... 48
Figure 15 Use case diagram of labour-neonatology-diaspensary process 49
Figure 16 HL7 RMIM for Labor Book design .................................. 50
Figure 17 Class diagram of the clinical main part of labor Book Application ................................................................. 51
Figure 18 Class diagram of Labour results part with relation to the Labor and Newborn classes .............................................. 52
Figure 19 Class diagram of relationship user-personnel-role .......... 52
Figure 20 Class diagram for export of statistic and medical information .... 53
Figure 21 Main window of labour book application .......................... 55
Figure 22 Example of application editor dialog windows ............... 55
Figure 23 Extended class diagram. Proposal of addition of classes Events, Record, Complex and Signal for storage of Partogram and Cardiotocogram

Figure 24 Events part of ontology ................................................................. 57

Figure 25 Flow chart of event option at creation ........................................ 60

Figure 26 a) The dividing of events in proposed experimental model, b) related dividing in HL7 ................................................................................. 60

Figure 27 Lead definition ........................................................................... 62

Figure 28 Process of analysis, derived signals, points and coplexes creation ............................................................................................................. 64

Figure 29 Relation of Events with the Characteristic points and Complex definition ............................................................................................................. 65

Figure 30 Class hierarchy and inferred class hierarchy for Lead of signal . 66

Figure 31 Signal part of ontology ................................................................. 68

Figure 32 Mapping between the HL7 aECG conceptual model and our model ................................................................................................................................. 70

Figure 33 Model of the ECG on the side of the physion of ECG reference ontology and signal part of our proposed ontology .................... 70

Figure 34 Basic model of signal evaluation ................................................ 73

Figure 35 The dependence of conditions to the type of signal ............... 74

Figure 36 Class diagram of Multi-layer model......................................... 76

Figure 37 Ontology of Multi-layer model .................................................. 77

Figure 38 Experiment, record, person, animal and role and part of ontology ......................................................................................................................... 80

Figure 39 Schema of basic arrangement of measurements in The Experimental Laboratory of Electrophysiology ................................. 81

Figure 40 Basic use case of initial processes of experiment ................. 82

Figure 41 Use case diagram of experiment ............................................. 82

Figure 42 Personnel part of ontology ....................................................... 83

Figure 43 Example of events of experiment in the ontology ................. 84

Figure 44 Example of event record application ..................................... 85

Figure 45 Example of signal part of experiment in ontology ............... 86
Figure 46 Class Hierarchy and Inferred Class Hierarchy for signal and its structures .................................................................87
Figure 47 Class Hierarchy and Inferred Class Hierarchy for complexes, interval and signal’s structures .........................................................87
Figure 48 Ontology graph of relations among signal, its structures, characteristic points and complexe and intervals.................................88
Figure 49 Signal reader interface with time axis creation .......................89
Figure 50 Scheme of signal processing, storage and visualization ........89
Figure 51 Example of Characteristic points and complexex interface with ECG signal and defined characteristic points and complexes...........91
Figure 52 Example of conditions, rules amd evaluation creation ..........92
Figure 53 Multi-layer model for data from experiments in The Experimental Laboratory of Physiology .........................................................94
List of Abbreviations

aECG the HL7 Annotated Electrocardiogram
AMSI American National Standards Institute
CDA Clinical Document Architecture
DASTA Data Standard
D-MIM Domain Information Model
DICOM Digital Imaging and Communications in Medicine
DTD Document Type Definition
ECG Electrocardiograph, Electrocardiogram
ECHR Electronic Healthcare Record
EEG Electroencephalograph, Electroencephalogram
EHR Electronic Health Record
EMR Electronic Medical Record
EPR Electronic Patient Record
EN, ENV European Standards
HDF HL7 Development Framework
HIS Hospital Information System
HL7 Health Level Seven
HMDs Hierarchal Message Definitions
ICD International Classification of Diseases
IEEE Institute of Electrical and Electronics Engineers
ISO International Organization for Standardization
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWL</td>
<td>Ontology Web Language</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RIM</td>
<td>Reference Information Model</td>
</tr>
<tr>
<td>R-MIM</td>
<td>Refined Message Information Model</td>
</tr>
<tr>
<td>SNOMED</td>
<td>Systematized Nomenclature of Medicine</td>
</tr>
<tr>
<td>UML</td>
<td>Unified Modeling Language</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
<tr>
<td>XSD</td>
<td>XML Schema Definition</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
1. Introduction

At a time when the development of IT and mobile technologies is precipitous, the need to communicate much more clearly is more and more important than ever before. While in the commercial areas of human activity is the need for a uniform data exchange format, information and knowledge is given clearly by financial motivation, in healthcare, this element is far more complex. Of course the financial motivation exists too, but it is only one of many. First we have to realize is that health care is still - especially in Europe - a very strong national domain. This is given by linguistic differences but also by different approaches to the question of interoperability (e. g. communication standard, whether architecture has a centralized point or is distributed). This situation appears particularly from the perspective of communication, exchange of data and information for which the interoperability and the standards are the primary assumptions.

Nevertheless, higher standards to ensure interoperability at the semantic and process level, bring another very important element. It is clear that the formalization of such a complex field such as medicine is an extremely complex task with large outputs. Standardization and unification of terminology in the form of nomenclatures, give rise to a very wide descriptive platform. With some advanced nomenclatures, such as e. g. SNOMED CT, it is possible to formalize not only the names of objects, organs, bodily processes, diagnosis or treatments, but it is possible to define relationships between them. With knowledge in areas such as knowledge representation, ontology, it is possible not only to transmit data and information but also their mutual relations. Having advanced
communication standards, such as HL7 standards, it is possible to define a large number of possible processes and their links to other objects or processes. We can say that the pursuit of interoperability, using standards and nomenclatures, allows representing knowledge in a structured, clearly defined form and keeping a large part of the context.

Let's look at this issue from the perspective of artificial intelligence. In addition to knowledge representation there is a very interesting task of data mining - whether the task classification, association or prediction. The process of data mining is already verified by a number of practical problems in various fields. In medicine it is often found in the form of clinical studies for pharmaceutical products or medical procedures. These studies are often prospective, that are very well prepared and managed. Obviously, if we had this contribution in retrospective studies, it would be very useful. But there we often encounter the problem of unstructured and non-complex data collection and storage. It is very difficult, if not impossible, to carry on such data relevant way of data mining or discovering of knowledge. Standards and nomenclatures to large extent make this opportunity possible. Moreover, if we are able to decide in the light of information and knowledge, the next step is to build a decision support system. If such a system is to be useful, its formalism must be properly designed and its advice must be presented in a clear and understandable form. Again, there are nomenclatures and standards that in conjunction with a suitable thesaurus (e. g. MeSH, UMLS) together allow fulfill these conditions.

1.1. Thesis contributions

The motivation of this thesis is to find the design and use of appropriate approaches to interoperability in medicine both for the purpose of communication and for knowledge representation, and not only for data, but also for their relationships and the treatment processes, for both categorical data and signals. Integration of the signal in the data model has been never solved in conjunction with draft of semantic description of signals (their behavior), and integration with other inputs of common information systems. This task is becoming increasingly important with the development of new technology and services. Crucial area in this regard is the
1. Introduction

telemedicine and monitoring of the health status of a patient in home environment. Also hospital departments require increasingly clinical data integration, mainly for the purpose of clinical research. This thesis has as the main aim the design of model, in accordance with the requirements of procedural and semantic interoperability in healthcare, in signals field linked events. This model is linked to existing solutions in fields of nomenclatures and standards in health care. To the best of our knowledge, we do not know about existence of such model/architecture. Its main part is the multi-layer model that contains part of events (acts), part of signals and part of evaluation representing background knowledge.

1.2. Thesis aims

The main aim of the thesis is proposal of an integrated data model, which allows description of all obtained data and analysis of data with minimal loss of semantic information. This approach should allow for a clearer linkage of technology (automatic) data processing and medical domains.

The thesis aims to develop description of the conditions and procedures that lead to the expansion of awareness and usability of collected data in a wider range. These procedures are demonstrated in two cases study from clinical and experimental part of medicine. Specific aims of this thesis are the following:

1. To propose an ontology model for heterogeneous data (clinical information and signal) with emphasis on the representation of the signal.
   • The ontology model defines relationships of classes in context of real word where the representation is formed by relation has or is. The model defines how there are formally linked subjects and observations within the architecture.
1. Introduction

2. To propose a multi-layer data model for heterogeneous data (clinical information and signal) with emphasis on the representation of the signal.
   - Besides the ontological model we define function structure of data model into three layers which represent single level of processing data from row form to complex description of signal. The complex description of signal represents the middle-level of knowledge.

3. To define this model in specific clinical area, specifically in the obstetrics area.
   - The first area of our efforts is clinical documentation. The reason for more complexity of record is given by ever-increasing demand for the possibility of processing data retrospectively. For such a process we need also raw data, such as origin signals.

4. To define this model in specific area of medical experiments, in particular electrophysiological and coronary experiments.
   - The experimental area is very useful for evaluation of proposed model because allows to use whole design.

1.3. Thesis outline

The thesis is divided into following parts:

- Chapter 2 deals with theoretical background of knowledge representation and ontology.
- In Chapter 3 and Chapter 4 there is presented the theoretical background from area of interoperability, standards and information technologies in medicine, design of systems and the state of the art in this area.
- Chapter 5 states the first case study in obstetrics area.
• Chapter 6 deals with extension of standards to the next level of data and information.

• Chapter 7 contains the second case study for extension approach of data collection in medical experiments.

• Chapter 8 contains summary of our proposal, main contribution and achievements of our work.

• Chapter 9 contains discussion and conclusion of the thesis.
2. Knowledge Representation and Ontology

Knowledge can be defined as a representation of human understanding of observations of the world. A goal of human endeavor to understand the observations are generalized models, and due to the subsequent use they can be defined as abstract models. A generalization of results of specific observation allows us to formalize relationships and implications of the observed phenomenon, process or state and thus represent knowledge. Knowledge representation is the cornerstone of the whole of human knowledge.

This chapter deals with the basic structure of the terms data, information and knowledge, and briefly explains how they are distinguished and how they are interrelated. The first part of the chapter is devoted to the basic division, the definition of knowledge and describes the distribution of knowledge. The second section summarizes possible approaches to knowledge representation, briefly stating their characteristics and use. The third part contains the definition and description of the most comprehensive approach to knowledge representation - ontology. This section briefly describes historical background, basic definitions, but mainly deals with the use of ontologies in computer science resp. artificial intelligence.
2.1. Data, Information and Knowledge

Terms data, information and knowledge are closely related together. Universally accepted definition for data and knowledge has not yet been accepted. One of frequently cited is view of G. Widerhold [1] in which knowledge and data are defined as complementary terms and distinction between data and knowledge can be expressed as follows:

“If we use automatic or non-expert human acquisition of material we talk about data. If we use experts for acquisition of material we talk about knowledge.”

Knowledge is abstraction and generalization of real world. It is usually less accurate and cannot be objectively verified. Data represent the current state of the world and contain many details, are large and often are changed. The knowledge deals with generalizations and does not change so often. Data and knowledge in the process of decision making [1] are shown in figure Figure 1.

![Figure 1 Function of Knowledge and Data in Enterprises](image-url)
According to [2] the relationship of data, information and knowledge is given as it is shown in Figure 2.

From the viewpoint of the use, knowledge can describe states as well as the processes. Then knowledge can be divided into:

- **Declarative knowledge** which represents knowledge about something
- **Procedural knowledge** which defines how to proceed in the implementation of some actions (inferring)

From the viewpoint of the accessibility, knowledge can be divided as:

- **Explicit knowledge** which is formalized and can be communicated and transmitted among humans and/or computers.
- **Implicit knowledge** which is hidden in the data, but which can be formalized by analysis of data.
- **Tacit knowledge** which is represented by hidden knowledge of experts and it is difficult to transfer this knowledge to another person.

In real world applications the experience shows that knowledge does not change so frequently as data. Thus new data acquisition may conflict with existing knowledge. However it does not necessarily mean that the knowledge is incorrect. In most situations we can find uncertainty both in knowledge and data. Therefore we have to introduce a measure of uncertainty that can express this fact as precise as possible. In some cases the conflict between knowledge and data may lead to introduction of new knowledge or modification of existing knowledge. In some cases updating
measure of uncertainty is sufficient. One of the interesting relation between knowledge and data is that data on higher level of processing or abstraction (aggregated or derived) can be viewed as knowledge at lower level of processing or abstraction. This is a fundamental issue that initiates consideration of multilevel systems and architectures. [3]

2.2. Knowledge Representations

Expert systems are systems defined in area of artificial intelligence, which were created as a tool for work on tasks with knowledge from specific area. These systems use to store of knowledge a base called base of knowledge. Knowledge is expressed in the base of knowledge by several different approaches. We will briefly describe knowledge representation by mathematical logic (propositional, predicate and description logic), rules, semantic networks and frames in the following section.

2.2.1. Propositional Logic

Propositional Logic is a relatively simple logic in which the only objects are propositions. The propositions are represented by propositional variables. The proposition variables can create sentences by combination of them with logical and the sentences can have more complex meaning. [4]

Language of propositional logic is created:

- propositional variables (e.g. X,Y,Z)
- logical constants
- logical connectors
  - \( \neg x \) means not \( x \)
  - \( x \land y \) means \( x \text{ and } y \)
  - \( x \lor y \) means \( x \text{ or } y \)
  - \( x \oplus y \) means \( x \text{ or } y \text{ but not both} \)
  - \( x \Rightarrow y \) means \( x \text{ implies } y \)
  - \( x \Leftrightarrow y \) means if and only if \( x \text{ implies } y \text{ and } y \text{ implies } x \)
- brackets as auxiliary symbols
2.2.2. Predicate Logic

This formal system is often used for the representation of sets and relationship of elements of the sets to each other and the subsets of sets. The expressions of predicate logic are formed from a number of elements and the combinations of elements extend the expressive power of propositional logic. Predicate Logic has logical connectors same like Propositional logic and two quantifiers, the existential $\exists$ ("there exists") and universal $\forall$ ("for all") quantifiers. [4] Language of predicate logic consists of:

- Term - variables and constants (for naming an object of the described world)
- predicate symbols (for defining relationships between objects),
- functional symbols (for defining functions)
- logical connectors $\neg$, $\land$, $\lor$, $\rightarrow$, $\leftrightarrow$ with the same meaning as in propositional logic
- quantifiers ($\forall$, $\exists$)

Language of predicate logic also allows to create formulas and the basic means of predicate logic expression are terms. Predicates are essential construct of predicate logic, where the predicates present one of the crucial parts of a sentence (describe characteristics of an objects). Particular relation is connected with number of predicates and thus the relation is a function of one or more arguments (with values true or false). [4]

2.2.3. Description Logic

Description logic is a mathematical formalism that ensures the formal platform for most of the ontology languages. Description is more expressive than propositional logic. The designation description logic is based on aims to describe concepts of world using a formal semantics based on mathematical logic. The language of description logic and predicate logic is identical in many elements and similar but not identical in using of notation to describe identical concepts (description logic has more efficient decision problems than predicate logic). Feature of description logic is to represent the structure of classes and relations as in systems based on frameworks or
ontologies, respectively. It is of particular importance in providing a logical formalism for ontologies. [4] There are several description logic languages:

- $\mathcal{AL}$ (Attributive Concept Language)
- $\mathcal{ALC}$ (extended Attributive Concept Language)
- $\mathcal{FL}$ (Frame based description language)

Detailed description of these languages can be found in [4], [5].

2.2.4. Rules

One of the most frequently used representations of knowledge are rules of IF-THEN type. In principle we can view the rules as another representation a situation of knowledge that can be represented formally by expression in mathematical logic. Since the rules can represent both static and dynamic knowledge they can be divided into two forms:

- procedural: 
  \[ \text{IF situation THEN action} \]
- declarative: 
  \[ \text{IF premise THEN conclusion} \]

Situation or premise and action or conclusions respectively are combination of statements about the current state of observed object. The procedural rules are common in monitoring expert systems (example is systems R1/XCON). The declarative rules are used in diagnostic expert systems (an example is system MYCIN). [3]

Example of rule from system MYCIN is shown in Figure 3.

![Figure 3 Example of a rule from MYCIN expert system [3]](image-url)
2. Knowledge representation and ontology

2.2.5. Semantic Networks

Semantic networks are more advanced knowledge representation then rules because they allow organizing knowledge around describe objects. Semantic network can be represented as oriented graph, where nodes are objects and binary relations between them are edges in the graph. The relations are essential for expressing knowledge. [3]

Semantic networks are modular because they can be easily extended by introduction of new object and relations. Example of semantic networks is shown in Figure 4.

![Semantic Network Example](image)

2.2.6. Frames

The frames are data structures, in which all knowledge about an object or event is stored in compact form. Although such representation cannot express more terms than predicate logic knowledge organization is suitable both for modularity and for access to knowledge. Moreover the frame system frequently allows specifying predefined values of information about an object when this information is not explicitly given.

Frames can easily represent knowledge on different abstract levels utilizing concept of inherence and structure of hierarchical class concept.
2. Knowledge representation and ontology

Knowledge representation by frames has served as inspiration for a type of programming languages. The appropriate style of these programming languages is called object-oriented programming because name for frames in this programming style is objects. [3] Example of frames (in Prolog syntax) is shown in Figure 5.

![Prolog code for a cylinder frame]

Figure 5 Example of frame

2.3. Ontology

Ontology is the explicit form of storage, formalized description of the given issue. Its basic structure is formed by ontology model objects, which are mostly consisted from the four basic parts: Individual, Class, Attribute and Relation. Ontologies are useful component in the fields of artificial intelligence, software engineering, biomedical informatics or Semantic Web as a form of knowledge representation about a part of the given domain.

According to the previous description, ontology is an ideal tool to achieve interoperability, mainly semantic interoperability. In the field of information technology and systems, there are a large number of implementations of ontology in various forms and for various purposes.
Ontologies have become common on the World-Wide Web (www). In [6] is presented Resource Description Framework which is developed in W3C Consortium [7]. The Resource Description Framework (RDF) is defined in [6] as a basis for metadata processing and it allows interoperable access between applications. Then the applications can exchange “machine understandable” information on the Web. The RDF can be used in many areas [6]: “in resource discovery to provide better search engine capabilities, in cataloging for describing the content and content relationships available at a particular Web site, page, or digital library, by intelligent software agents to facilitate knowledge sharing and exchange, in content rating, in describing collections,…” (see more in [6]). The description of RDF can be found also in [4].

RDF present data model used by the Semantic web and OWL (Ontology Web Language). OWL was developed by the W3C [7] as a general tool for the Semantic Web. OWL is very useful tool but also relatively complex language. OWL can be presented as an extension of RDF Schema. RDF schema is a formal definition of the syntax of a language, and a schema language is a language for expressing that definition (like XSD to XML). OWL compared to RDFS allows performing more powerful inference. [4]

Ontologies are developed in many areas where knowledge in given domain can be used to share and annotate information in these areas. There is vocabulary for ontology creator for share information that contains machine-interpretable definitions of basic concepts in the given domain and theirs mutual relations. One of the most important aims of ontologies developing is sharing understanding of the information structure among people or software agents. [8]

There are several reasons for creation ontology [8]:

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge
Similarly to the knowledge, ontology has not, in the field of artificial intelligence, a generally accepted definition. We will use the definition given in [8], because we use Protégé [9] application for creation of our proposed ontology. According to this definition:

“Ontology is a formal explicit description of concepts in a domain of discourse (classes or concepts), properties of each concept describing various features and attributes of the concept (slots or properties), and restrictions on slots (facets or restrictions).” [8]

Ontology together with a set of individual instances of classes constitutes a knowledge base. In practical terms, developing ontology includes [8]:

- defining classes in the ontology,
- arranging the classes in a taxonomic (subclass–superclass) hierarchy,
- defining slots and describing allowed values for these slots,
- filling in the values for slots for instances.

The classes are representation of domain model (concepts) and are mostly the main focus of ontologies. Ontology is a model of real state of the domain and the concepts in the ontology have to reflect this state. And we want evaluate and test the knowledge. It can be performed by using it in applications or solving methods or by discussing it with experts, or both. [8] Example of simple ontology created in Protégé [9] is shown in Figure 6.

![Figure 6 Simple example of Ontology - ontology of wine](image)
In Protégé are defined for using of properties and classes simple language which is based on mathematical logic (viz. sections 2.2.1, 2.2.2, 2.2.3). The definition of this language is following [10]:

- some – an existential restriction
- only – a universally quantified restriction
- min – minimum cardinality restriction
- max – maximum cardinality restriction
- exactly – an exact cardinality restriction

Example of the possibilities of using ontologies for integration and storage of data and reaching semantic interoperability is the article [11]. This is a detailed description of the possibilities of using the ontology as an integrator of multiple sources of information with links to biomedical applications, in which number of different sources is expected. Ontologies are important in biomedical research through a variety of application. Ontologies are used mainly as a source of code books and vocabulary for standardization and integration purpose. Ontologies are also used as a source of computable knowledge. Typical examples of biomedical ontologies in knowledge management, data integration and decision support is presented in [12]. Using ontology in information system in dental clinic is presented in [13]. Ontology for EEG/ERP Domain can be found in [14]. In [4] are presented several Bio-Ontologies (Gene Ontology, Cell Ontology, Ontologies for Anatomy, etc.).
3. Interoperability, Standards and Information Systems in Medicine

This chapter introduces three basic areas important for the following parts of the thesis, namely interoperability, standards, and electronic health record. These three areas are interconnected and give basic direction.

3.1. Interoperability

In this section we deal with the concept of interoperability, its definition and possible division by the level of exposure. There can be found several definitions of the term interoperability. Here there are two slightly different ones that clearly express the meaning of interoperability.

The frequently used definition is from the IEEE Standard Computer Dictionary [15]:

“Interoperability is ability of two or more systems or components to exchange information and to use the information that has been exchanged.”

Another one is the definition from [16] more focused on computer science:

“Being able to accomplish end-user applications using different types of computer systems, operating systems, and application software, interconnected by different types of local and wide area networks.”

These two definitions clearly illustrate that the concept of interoperability is very significant in today's world. Almost every area of
human activity has reached such a level and scale that it is not possible for them not having defined formalisms, at least in the assumption of their mutual interaction.

Interoperability can be divided into several types, according to which level it reaches interaction of systems. Here we provide a breakdown and description of interoperability according to work [17], where interoperability is divided into three levels:

- Technical interoperability
- Semantic interoperability
- Process interoperability

3.1.1. Technical / syntactic interoperability

Technical interoperability means the ability of two or more system to exchange data with each other regardless of the domain, and no effect of distance. In [18] there is one of the foundations of technical interoperability using Shannon's information theory [19] indicating the possibility of achieving 100% reliability of communication over a noisy channel. However, the technical interoperability of systems does not know or care about the meaning of the transmitted data. The question is whether the level of technical interoperability can be merged with the standard defined as syntactic interoperability. Both have in common that they are exchanging data without knowledge and concern for their meaning. But for syntactic interoperability there is more accurate specification using the format of data or communication protocol. Typical example at a lower level interoperability is data encoding formats such as alphabetical characters in ASCII or other format, or at a higher level such as XML and SQL.

3.1.2. Semantic interoperability

According to [17] and [18] semantic interoperability ensures that all given systems understand exchanged data – it means that the computer can interpret and use the information in context and without ambiguity. The reason for ensuring of semantic interoperability is achievement of the
usefulness of exchanged information and more powerful using of decision support systems in healthcare area. Semantic interoperability is specific into given domain and context and it is usually reached using of codes and identifiers.

According to EN 13606 Association [20] semantic interoperability is the ability of two or more computer systems to exchange information, semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems.

“To achieve semantic interoperability, both sides must refer to a common information exchange reference model. The content of the information exchange requests are unambiguously defined: what is sent is the same as what is understood.

In the field of health information, to achieve semantic interoperability is even a more important and difficult duty. The complexity of the health domain, its frequent variation and evolution and the differences between the information technologies domain and the health domain need a deep change on the methodologies of information management.” [20]

Semantics of communication in system or among several systems are mostly results of compromise between various requests to syntax versus pragmatics of system. It is similar to combination of software specification that is often compromise from stepwise design (defined mainly from user requirements) and bottom-up approach.

3.1.3. Process interoperability

We proceed from the work [18], where process interoperability is defined as the coordination of business processes allowing different systems of different organizations to work together. The condition for functional interoperability is that participants of processes share understanding about interoperability of business systems and there is coordination of work process. Benefits are achieved only when the systems are used in everyday work.
3.2. Standards and nomenclatures in medicine

Interoperability is transferred to the specific form by creation of standards in the given field. In this section, we will briefly deal with several important standards in health care and in the transfer of information, particularly those describing the health of the patient. These standards are created mostly by international organizations active in the field of standardization (e.g. the health informatics). These standards are then taken over to the national standards, thus ensuring their effectiveness, e. g. in the Czech Republic standards are classified by one of three ways: translation, original acceptance or endorsement.

Standards are the basis for interoperability among different implementations of EHR. Standardization can be divided into 4 levels [13]:

- Structure - specification of communication, its format, sorting, validation, a typical example is the specification of the structure of XML messages using a file using DTD documents or XML schemas.
- Content - definition of terminology and code system (nomenclatures) designed for the uniqueness of the transmitted data, their coherence and meaning.
- Technology - definition of possible tools for storing and exchanging information.
- Organization - definition of processes and procedures as clearly reproduce and combine them with other information, e. g. defined medical procedures in the Czech Republic.

For more clarity between structural and semantic standardization we are going to describe these two types of standardization separately.

3.2.1. Structural Standardization

All mentioned standards are related to electronic health record (EHR). EHR is an essential component of the hospital information system (HIS). Its
mission is to collect and store health information about the patient such as diagnoses and medical history, information about their investigations and observations, information on the performed treatment (medication, surgery, and therapy). EHR is a critical part of the information system in healthcare facilities and it is essential for the patient's medical records. Nowadays, when a healthcare is increasingly distributed, it is necessary to define how the exchange and sharing of health data will be carried out. [13]

Standardization should be considered as the basis for designing architecture that defines methods of communication and data representation. In general, at present the standards are mostly used at a local level. The reasons do not arise only from the technical difficulty of the problem, but also from the legal and political rules which represent important part of the problem. Another big issue is already a great variety of existing proprietary or to same degree standardized interfaces that should allow EHR integration in heterogeneous systems. We will focus on the most important standards (in our opinion) for the exchange of reports from medical records: Health Level 7 (HL7), ENV 13606 or openEHR and DICOM.

3.2.2. Content Standardization

Besides the standards for communication, standards defining the content of medical records are very important. Many of the data is clearly defined and linked. It is appropriate for these data to store them in structured form with hierarchical classification, which allows not only a clear exchange of information, but also their clear inclusion and importance. The next sections briefly describe the three major code systems used in most communication standards for systematic (categorical) data in medicine. They are the International Classification of Diseases version 10 (ICD-10, ICD-10) defined by the World Health Organization (WHO), the Systematized Nomenclature of Medicine (SNOMED) and laboratory coding system Logical Observation Identifiers Names and Codes (LOINC).
3.2.3. HL7 standard

HL7 is a nonprofit organization that develops standards in healthcare. HL7 is registered under the American National Standards Institute (ANSI) [21]. HL7 is organized as international organization based in the United States and affiliate organizations in many countries. The exact distribution of the HL7 organization can be found in [22].

HL7 develops a wide range of standards in health care both in the clinical area, the economic area and management area. Now HL7 version 2.x is widely used and established standard of HL7 for data exchange. This standard defines the format and content of reports to communicate and transfer data among various participants of the health care processes (exchange of information about patient state, results of measurements, etc.). However, HL7 version 2.x has limitations in insufficient definition and validation of message formats. Therefore it is developed HL7 version 3 that allows writing of messages using XML format and thus using object-oriented approach and thus more complexity of messages. Reference Information Model (RIM) is the basis of HL7 version 3. RIM is an abstract general model of closed universe that is defined using classes, their attributes and relationships between classes. HL7 RIM core classes, basis attributes and relationship between classes is shown in Figure 7 [23]. To specify classes, attributes and relationships in a particular domain of the given area (clinical medicine areas, economy, and management), another model is derived from RIM. This model is called the Domain Information Model (D-MIM). The derivation of D-MIM from RIM is carried out by two basic operations - the specifications and restrictions of RIM classes, theirs attributes and relationships between classes. The next step to specific model with information about the patient's medical record is derivation of the Refined Message Information Model (R-MIM) from D-MIM. R-MIM is a subset of D-MIM. This model has primary graphic expression, but has also tabular form of information content of given abstract structures of messages that are called Hierarchical Message Definitions (HMDs). Based on R-MIM or HMDs abstract Message Types (MT) are created. Adding data into abstract structures of MT creates a specific message. HL7 version 3 is useful, very strong and comprehensive approach for both modeling
processes in medicine and communication among medical information systems. But just the complexity increases the difficulty in both proposal and implementation of systems in accordance with this standard. This is one of the main reasons why the HL7 version 3 is not as widespread and the HL7 version 2.x implementations are more frequently used. [13] [18] [22]

Clinical Document Architecture (CDA) [24], [18] is the most often used part of HL7 version 3. CDA is a document that contains a header and body. These two parts are related to HL7 RIM and rate of the relation is given by level of CDA. CDA is divided into three levels [18]:

- CDA Level 1 has a header and human-readable body. The header contains basic meta-data (for retrieval of information) and body is human-readable text or image (text document like PDF, image like jpg).
- CDA Level 2 allows the body to be either an unstructured blob or one or more structured section. Each section contains a single narrative block, which contains XML markup that can be rendered in human-readable form.
- CDA Level 3 allows each section to include machine-processed entries at almost any level of granularity. Thus, it offers the benefits
of both human-readable and machine-processed documents. Machine-processed data are encoded using the HL7 version 3 Clinical Statement pattern. However, CDA Level 3 is now not available and it is only in proposal form.

The detailed description of CDA and CDA, Release 2 are in [25], [26], [24]. We mention several works that are most relevant to our work in the field implementation of HL7 standard, mainly in version 3. First of all, it is work [27]. This work describes architectural approach for implementation of HL7 v3 Information System. This work is especially important, because it clearly shows how the development of the entire field should be directed. It is primarily the creation of architecture for a given area and subsequently the implementation of the system. And secondly, we present Figure 8 from [27]. The Figure clearly reveals the essence of HL7 communication. HL7 standard allows to work in a distributed system environment (Figure 8a) and allows a much wider range of possibilities, not only in the exchange of information between healthcare facilities, but also in the field of public health. Therefore, it is not necessary to create a central point that would bring together all data. However, if this element is desired, it can be created using the HL7 server. The HL7 server has clearly specified communication conditions, transformation of messages and filters (Figure 8b). An example of HL7 server is open source project Mirth Project [28].

Figure 8 a) Point to Point HL7 interface, b) HL7 Messages Server [27]
Linking multi-agent technology, ontologies and HL7 standard is described in the work [29]. It is a system in the role of facilitator that manages the flow of patient information across a healthcare organization. Thus, connection of agent technology and standard enables the creation of robust tools for unique integration of the required information from many different sources. In [30] there is described an approach to sharing data and knowledge using HL7 RIM. This work presents Unified Service Action Model (USAM) as part of the HL7 RIM that allows conceptual integration of patient information and medical knowledge. The USAM part of the entire RIM contains act relationships for medication, observation, procedure, condition node and transport. This concept is important for our proposal model and is related mainly with the event part of our model. Finally, let us mention the paper [31], where Autonomous mapping of HL7 RIM and relation database schema is described. This paper is interesting for us for two reasons. Firstly, the RSM architecture supposes mapping knowledge repository. Secondly, HL7 JavaSIG Framework is used for mapping of RIM based custom classes and healthcare database which is connected over Hibernate Mapping file [32]. The HL7 JavaSIG Framework (http://aurora.regenstrief.org/javasig) is an API for message generation and parsing. The message is parsed into the corresponding custom class and stored into the database using hibernate mapping file. Hibernate [32] is an open source framework that persists objects in relational database transparently, very well matured and adopted by a large developer community.

3.2.4. CEN TC 251 a ENV 13606

CEN TC 251 WG 1 issued European standard EN 13606 Electronic health record communication [20].

The standard defines two important models; The Reference Model:

“A Reference Model is an Object Oriented model that is used to represent the generic and stable properties of health record information. It comprises a small set of classes that define the generic building blocks to construct EHRs. Typically, a Reference Model contains:
A set of primitive types.

A set of classes that define the building blocks of EHRs. Any annotation in an EHR must be an instance of one of these classes; we will call them entities. For instance EN13606 defines six different types of entities: folder, composition, section, entry, cluster and a set of auxiliary classes that describe the context information to be attached to an EHR annotation including versioning information.

It may contain classes to describe demographic data and to communicate EHR fragments.” (http://www.en13606.org/the-ceniso-en13606-standard/reference-model) [20]

And Archetype Model:

“An archetype is a structured and constrained combination of entities of a Reference Model that represents a particular clinical concept, such as a blood pressure measurement or a laboratory analysis result. This structure should be defined by a health domain expert. An archetype can be also defined by further constraining another archetype, called parent archetype, in order to obtain a more adequate or fine grained representation of the clinical concept.

Archetypes are composed of three main sections: header, definition and ontology. The header section basically contains metadata about the archetype, such as an identifier or authoring information. In the definition section the information is where the clinical concept, which the archetype represents, is described in terms of Reference Model entities. Finally, the ontology section is where the entities defined in the definition section are described and bound to terminologies.” (http://www.en13606.org/the-ceniso-en13606-standard/archetype-model) [20]

3.2.5. DICOM

Digital Imaging and Communications in Medicine (DICOM) [33] is an international standard (ISO 12052) that defines formats for exchange of medical images from several imaging modalities (e.g. X-ray, CT, MRI,
ultrasound, etc.). DICOM is one of the most widely and commonly used standards in world.

DICOM Structured Reporting (DICOM SR) [34] is based on DICOM standard, which is extended with structured hierarchical document of medical reports. Thus DICOM SR can be used as patient overview data exchange standard. On the semantic level data exchange is ensured by using of some of the coding systems (see ICD, SNOMED, LOINC). [34] [13]

3.2.6. ICD 10

For description of the diagnostic standard we use The International Classification of Diseases (ICD) [35] published on the official website of the World Health Organization (WHO) [36].

“ICD is the standard diagnostic tool for epidemiology, health management and clinical purposes. This includes the analysis of the general health situation of population groups. It is used to monitor the incidence and prevalence of diseases and other health problems. It is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States. It is used for reimbursement and resource allocation decision-making by countries.” [35]

3.2.7. SNOMED

For description of terminology we use SNOMED. We proceed from the official website of the International Health Terminology Standard Development Organization [37], owner of standard SNOMED Clinical Terms (SNOMED CT):

“SNOMED CT contributes to the improvement of patient care by underpinning the development of Electronic Health Records that record clinical information in ways that enable meaning-based retrieval. This
provides effective access to information required for decision support and consistent reporting and analysis. Patients benefit from the use of SNOMED CT because it improves the recording of EHR information and facilitates better communication, leading to improvements in the quality of care.” [37]

3.2.8. LOINC

The Logical Observation Identifier Names and Codes (LOINC®) [38] database provides a universal code system for reporting laboratory and other clinical observations. Its aim is to coding of observations in electronic messages such as Health Level Seven (HL7) messages. Based on the coding then hospitals, health maintenance organizations, pharmaceutical manufacturers, researchers, and public health departments receive can exchange messages from various sources and automatically sort exchanged data in the right slots of their medical records, research, and/or public health systems. LOINIC is included in the HL7 standard as laboratory dictionary, but its use is limited mainly to the United States. In the Czech Republic there is a laboratory standard called National Standard of Laboratory Items in the national data standard [39].

3.2.9. openEHR

The openEHR [40], [41] approach to modeling information, services and domain knowledge is based on a number of design principles. The application of these principles lead to a separation of the models of the openEHR architecture, and consequently, a high level of componentization. This leads to better maintainability, extensibility, and flexible deployment.

3.2.10. Continua Alliance

Continua Alliance [42] is a nonprofit coalition of companies from the area of healthcare and information technologies that aims to offer a solution for interoperability at the level of personal health. This solution is mainly in the field of telemedicine, home treatment and care, empower individuals. Continua wants to enable use home monitoring of basic physiologic
parameters like weight, blood pressures, glucose, etc. and share it with other healthcare information. [42]

Architecture of Continua is divided into four levels with three communication channels. The levels are Personal devices, Aggregation manager, Telehealth service center (is optional) and Health Records. The communication channels are device connection, wide area network (WAN) connection and health record network connection. The Connections are standardized using common standards like Bluetooth, USB, ZigBee, ISO, IEEE, wifi, W3C and HL7. [43]

3.2.11. IHE

“Integrating the Healthcare Enterprise (IHE, http://www.ihe.net/) is an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information. IHE promotes the coordinated use of established standards such as DICOM and HL7 to address specific clinical need in support of optimal patient care. Systems developed in accordance with IHE communicate with one another better, are easier to implement, and enable care providers to use information more effectively.” [44]

IHE was founded in 1999 by the Healthcare Information Systems and Management Society (HIMSS) and the radiological society of North America (RSNA). The aim of IHE is to improve way of communication, exchange and sharing of information among healthcare computer systems (mainly using HL7 and DICOM standards). IHE does not participate directly in the development of the standards, but defines condition of their using to achieve the higher level of interoperability. Systems developed in accordance with IHE profiles ensure better communication and exchanging among systems and have clearer implementation. IHE has used a four approach of process of collaboration and communication among key parties [18]: Identify Interoperability Problems, Specify Integration Profiles, Test Systems at the Connectathon and Publish Integration Statements for use in Request for Proposals (RFPs). [18] [44]
3.3. Information systems in medicine

At the end of this chapter, we would like to come back to the notion of Electronic Health Record (EHR). Terminology in electronic health record is not uniform and is also often used incorrectly. They are often confused concepts of automated medical record, digitized medical record, electronic medical record, electronic patient record and electronic health record. Therefore, we want to clarify these concepts. We start from the following description of the book [23] and U. S. Medical Record Institute. Single record forms differ both in terms of content and level of digitization. Differences among single record forms are shown in Figure 9 and differences among Electronic Patient Record, Electronic Health Record and Electronic Healthcare Record are described in the following points [23].

- Automated Medical Record is a paper based medical record with some automation in medical documentation.
- Computerized Medical Record is created by digitalization (scanning) of medical record documents and storing digital file. The structure and view of documentation is same like paper record. The computerized medical record does not use OCR or ICR text recovery function, it is pure image system.
- Electronic Medical Record (EMR) is a digital medical record including data management. EMR is embedded in IT-based
organization support of clinical processes like decision support and interactive guidelines. EMR is connected with business and hospital management data.

- **Electronic Patient Record (EPR)** is a system that contains all disease relevant data of a patient. It can be established beyond an institution and exceeds the framework duty within a medical record. Nowadays, the term electronic patient record has been largely replaced by the broader concept of the electronic health record.

- **Electronic Health Record (EHR)** is an extension of the electronic patient record, in which apart from the care of a particular patient there are also taken into account education, science and research. There is a standard called universal health record, which specifies the requirements for clinical record maintained in electronic form. These requirements are divided into four basic levels - architecture, terminology, communication, and security.

- **Electronic Healthcare Record (ECHR)** is a form of digital storage (mostly databases) of complete medical records. It is integrated electronic health record data with the data of administrative character that provides hospital information system.

Hospital Information System (HIS) is a system consisting of several modules for different areas that are processed in hospital facilities (e.g. clinical or outpatient module, economic module, management module, prescription module or also library module, etc.). The main objective of HIS is to provide comprehensive information services for hospital equipment. For medical part of HIS in electronic form an electronic medical record is the core of system. The HIS is proprietary system and can be customized for particular hospital. Communication in system or among the systems can be based on either paper or can be paperless (electronic) or can be combination of both approaches. In electronic form the communication and exchanging issue is solved as one of the functions of the electronic system. [23] [13]
3.4. UML and Design of Application

We proceed in this subchapter from book [18]. HL7 Development Framework (HDF) is documentation of development methodology for HL7 standards. The HDF uses project oriented method, which contains two basic components - Product Life Cycle and Product Development. The HDF has five main phases [18]:

- Project Initiation Process
- Domain Analysis Process
- Specification Design Process
- Standard Profiling Process
- Technology Specification Process

Analysis process contains important part of proposal: Domain analysis, business context, use case analysis, process flow, information model and business trigger analysis. HDF methodology uses as tool mainly unified modeling language. [18]

Unified Modeling Language (UML) is a specialized and standardized general purpose modeling language, for writing of structured information from hierarchical messages to a notation for drawing diagrams. UML was firstly used in the list of OMG adopted technologies in 1997, and since it is used in many standards for modeling of software systems. UML contains system of notation that is used on diagrams and meta-model. UML diagrams contain either information structure of some entities or behavior of some entities. UML contains up to 13 diagram types. [18] Let us mention here a brief overview of UML diagram types.

We use mainly use case diagram and class diagram for description of our proposal. The use case diagrams are used for description of behavioral requirements of business processes. The systems in our work are firstly modeled by use case for description of all aspects of systems. The class diagrams are the most frequently used UML diagrams. They show the structure of classes with definitions and attributes, and of course relationships of classes between them. [18] We use class diagram to specific final parts of our systems.
3.5. XML data format and XML schema

XML (eXtensible Markup Language) is a markup, general format for creation of documents based on structured data [18]. XML defines a set of rules for encoding documents in both human-readable and machine-readable forms and is used as interoperability means between different systems or applications. XML documents are independent on the application that create or use them. [45]

The W3C XML Schema Definition Language is an XML language for definition of structure and constraining of the XML documents content. W3C XML Schema is a W3C Recommendation. In [46] there is an introduction to using W3C XML Schemas, and this contribution also includes a comprehensive reference to the Schema data types and structures.

The creation of schemas is issue of the analysis and design for creating and changes of XML document. Schemas can be created for use with existing sources. If uniformity of document structure is necessary then can be specified strict rules for structure and content of document. The schemas make the rules explicit and XML allows tagging of data. [18]

XML is the communication basis of HL7 v3 on the lowest level. In the end of the modeling process there are the Refined Message Information Model (RMIM) and Hierarchical Message Description (HMD). HMD is only tabular format expression of RMIM and both contain same information (RMIM is graphic form, HDM in tabular form). RMIM or HMD are basis for creation of related set of message type. A message type (MT) is a specification of an abstract message structured, which can be completed by particular data and thus create particular message for interchange. Finally, MT is send as a linear string of XML and validate using XML schema. [18] Example of HL7 message in XML format is shown in Figure 10.
3. Interoperability, Standards and Information Systems in Medicine

Figure 10 Example of HL7 V3 XML message
4. Processing of heterogeneous data

Heterogeneous data is defined as data originating from several different sources stored usually in different formats. In evidence based medicine, the decision making process nearly always involves dealing with heterogeneous data. However, automatic processes handling and evaluating such data are still very much limited in the sense of full automatic heterogeneous data processing and in ability of representing the relations among the data. In medicine defining the relations between data from different sources is an integral part of the cognitive process. But the volume and heterogeneity of the acquired data still pose great problem on manual and classical automatic (defined on one constraint field of data sources) processing of the data. This current state of handling heterogeneous data and information in the medical domain motivates us to try to find methodology of the processing of heterogeneous data in general. We focus on approaches for heterogeneous data storage, fusion and processing, because these approaches represent a very important part of our background for the development of a multi-layer data model.

In the knowledge and data engineering area the heterogeneous data is often represented by a data model, which is interchangeable and suggestible and represents information of each source and makes possible combining such information. A goal is to acquire new knowledge from the inner relations in the data from several sources. Representation used for integration of heterogeneous data needs to be competent in data description, schemes and contexts. For the integration purpose different approaches had been used such as object oriented models [47], ontology [48], [49], statistical model [50] or mediated data model [51]. In [48] there is a survey of generally recognized the main heterogeneity integration conflict.
4. Processing of heterogeneous data

categories – syntactic, structural and semantic heterogeneity. Approach to semantic heterogeneity is very important for interoperability limitation. The ontology approach offers solution of this limitation [48]. The role of ontologies, metadata and context in information systems semantic heterogeneity is introduced in [52]. Overview of main ontology architectures (single ontology approaches, multiple ontologies, hybrid approaches) and ontology representations can be found in [49]. For ontology visualization there exist several approaches and a number of techniques used in other contexts that could be adapted for ontology representation [53].

Data fusion is one of the most important parts of heterogeneous data processing. Data fusion is concerned with combining information from knowledge sources such as sensors to provide a greater understanding of a given situation. In paper [54] there is described the fusion of ECG, blood pressure, saturated oxygen content and respiratory data for improving clinical diagnosis in critical care unit. Many parameters are derived from original record of physiologic parameters (i. e. heart rate, respiratory rate, systolic, mean and diastolic pressure) and these parameters are used as features to data fusion. Data fusion is represented by Rule based approach and Fuzzy based approach. Paper [55] deals with concept of learning rules from multisource data from monitoring devices in Cardiac Intensive Care Units. Multi-class Inductive Logic Programming is used to learn the rules in two steps. In the first step rules are learned independently from each source. In the second step the learned rules are used to bias a new learning process from the aggregated data. Aggregation is the operation consisting in merging examples from different views of the same situations, which is dependent on learning data and can be different from one multi-source learning problem to another. In paper [56] there is described observation of Ventricular Rhythm based on ECG or hemodynamic response (blood pressure). Work [57] presents a system for evolutionary induction of decision trees and automatic programming (AREX), which is used for the discovery of a possible new medical knowledge in the pediatric cardiology area. All these works have very strong emphasis on the importance of role of rules in the field of biomedicine.
Main attribute for setup of heterogeneous data is to assess affinity evaluation, which exists among elements (attributes). To express the affinity evaluation heterogeneous distance functions is usually used. More extensive survey can be found in [58]. In [59] global affinity is described by three parameters, which express level of affinity between elements for obtaining global view of heterogeneous data sources. In [60] statistics distance-based model is used for complete rankings of heterogeneous data. Application of heterogeneous value difference metric for elements as affinity measure is used in [61].

For the analysis of relations of the data from different sources it is possible to use the artificial intelligence methods. In general all data-mining techniques can be considered. Overview of such methods is to be found in [62]. One of the main tasks in the area of dealing with heterogeneous data is to assess their similarity based on the above mentioned metrics. For this purpose clustering is often used. In [59] hierarchical clustering technique of agglomerative type [63] is used. Another interesting example of clustering algorithm is in [60], where the clustering task is solved using expectation-maximization (EM) algorithm [64].

Not only clustering but also classification can be used to deal with the data. In [61] artificial immune system method [65] is used to classify heterogeneous data. On the other hand [50] uses dependence trees method for classification. Of course also when dealing with heterogeneous data the dimensionality problem (REF) is an obstacle to be tackled. One approach to the data dimension reduction problem is shown e. g. in [66]. Issues of classifier training are main theme of [66], [67]. Article [68] provides a comprehensive evaluation of a set of diverse machine learning schemes on a number of biomedical datasets.

In the area of medical problems heterogeneous data is frequent. Medical diagnostics is based on assembling formations from various sources – patient and family history, basic patient data (age, sex, height, weight, etc.), signal measurement (ECG, blood pressure, respiration recording) and their analysis, information from imaging systems (ultrasound, CT, MRI) or results of laboratory analysis - with final diagnosis if available. Often only one source is used for automatic classification. There are some papers combining several sources e. g. [56] uses fusion of electrophysiological and
4. Processing of heterogeneous data

hemodynamic signals for ventricular rhythm tracking. In the area of electrophysiology, e. g. in [69], body surface potential mapping (BSPM) and MRI-based model for solving of inverse task [70] are used. However, in those two cases relation between sources is a priori known.

For dealing with heterogeneous data in general this approach is not sufficient. One of the common cases for using heterogeneous data is in risk stratification. As an example work by [71] can be used, where the artificial intelligence methods are used for intelligent analysis in predicting outcome of out-of-hospital cardiac arrest. Data mining in medical area is very difficult process. Uniqueness of medical data mining is in more detail described in [72]. The major points of uniqueness of medical data are given in four headings - heterogeneity of medical data; ethical, legal and social issues; statistical philosophy, and special status of medicine [72]. Overview of used technologies for data mining process model in [73] is introduced and examples of medical data mining cases with various approaches can be found in [74], [75], [76], [77]. Process of data mining and its results could be very difficult and large for classical visualization. Therefore there are used improved and special methods of visualization. Overview of data mining task, data, models and results visualization methods is presented in [78]. In [79] clinical decision support system architecture is presented.

The paper [80] is very important for semantic interoperability in signal area. The paper presents ECG reference ontology for semantic interoperability of ECG data. Ontology is based on decomposition of ECG into wave (P, Q, R, S, T), complex (QRS) and segments (e. g. PR, ST, TP). This approach is essential because only in this way we can achieve a complex signal description in semantic form. This paper also presents a mapping of reference ECG ontology to three commonly used standards of ECG storage: AHA/MIT-BIH, SPC-ECG conceptual model and HL7 aECG. The results of this paper are important for our work and we have expanded the idea of article in more general considerations and wider links to other type of data.
5. The First Case Study: Obstetrics Area

In this chapter we describe an ontology model of delivery domain and a design of clinical system in obstetrics, which is based on the ontology model. The purpose of this chapter is to show the possibility of designing and developing a system in a specific real healthcare area with all its obstacles. At the beginning it is necessary to mention that the scope of the design of this work is limited by the requirements and possibilities of the contracting authority. Given shortcomings of systems, given especially by the date of their implementation, there is often a problem with the transfer of data. Note that the design and developing of the system has been done in Czech environment, thus respecting the standards used in the Czech health care. Systems enable communication using the Czech National Data Standard [39], but however this standard is mainly used for the transmission of data to state authority - Institute of Health Information and Statistics (UZIS, http://www.uzis.cz/). Usually, these systems have not proper interface implemented (or its definition), which would allow to obtain anonymised data that could be used to create data sets to more complex data analysis. The systems do not allow communicate and exchange information with other systems via clearly defined interfaces. The consequence is that the information contained in one system is not easily accessible (or even inaccessible) for other systems. Another issue is when we need to enter the same information more than one and the forms are not property structured.
The result is that the same information can be stored in different ways and the queries do not return the expected records.

Our proposal focuses on the clinical part of the process and is designed mainly for the acquisition of medical information. In this case, the necessary information is the information about mother before labor, during labor and subsequent information about the newborn and his/her eventual hospitalization and treatment. Relationship of delivery and neonatology is important primarily due to the need to provide information for subsequent dispensary of the child, especially if that is related to complications during labor.

5.1. Ontology of Obstetrics Domain

As a basis for subsequent work we define ontology as a model of reality in the area of obstetrics. An application design is based on this ontology model. We have used Protégé [9] for creating of our proposed ontology. A basis is defined by the following set of classes:

- Labor
- Mother
- Newborn
- Neonatology
- Parameters
- Process

Each of the classes will be described and analyzed in the design of application as well as their interactions. Ontology is defined in this section as the relationships in a logical structure of classes. Ontology defines relationships: a class has a subclass. Therefore, the definition from Figure 11 cannot be taken as the definition of the functional interface, but as a hierarchical description of the semantics of the issue. Class Labor has primarily as subclass its parameters indicating the facts that they relate directly to it (the date and time, a certain medication, etc. ). Next important subclass of Labor class is a subclass Process. It shows how the process of
labor is evaluated and it usually correlates with the class ResultOfLabor. Class Mother has to type of value with respect to delivery outcome. Firstly, class Mother has several defined values that are important for the evaluation of delivery and its outcome (e. g. gravidity, parity). Secondly, class Mother includes data that are important for medical documentation as a whole, but not in respect to delivery (e. g. residence). Class Newborn has, besides some basic information about a newborn, subclass Neonatology. It contains information about the newborn after passing the newborn from a department of obstetrics to a department of neonatology and then to care of pediatricians.

Further, the ontology defines some additional classes that are not defined in labor book application, but they are important in terms of the completeness of the model. It is the class ExpectantMother, which is associated with a class of Pregnancy. It is particularly important to realize that there is a record of the course of pregnancy, and if it is requested, it is possible to connect this record to the record of the delivery. The same holds for the Dispensarisation. These classes are not implemented in the application, but it is important to have this part ready and be able to add the dispensary record to the record of the delivery. The dispensary has an area, in which it is implemented, and diagnoses, which are defined at discharge of the newborn from neonatology department.

The most important issue is presentation of the concept of events that defines the states, conditions and processes that take place. The events are defined either by the user (expected or possible) or events are subsequently found and defined based on analysis. In our proposal, events are divided into diagnosis, intervention, medication, measurements and events. Events of diagnosis type are clearly defined. Medication event may be further divided by type of medication, such as a bolus, infusion, etc. Event of intervention type may include any event that intentionally interferes with the organism's natural activities, except medication. Difference between an observation and a measurement is important. As a measurement, we call every event that is recorded using the device and we can take it as objective. On the contrary, the observation is an event based on subjective observation and assessment of human (expert). We describe the events further in the next chapter.
5. The First Case Study: Obstetrics Area

Figure 11 Ontology of Labour
5.2. Assumptions of the system

The purpose application presented in next sections of this chapter is to create an electronic form of document, which is called labor book, to store records in a single, full-form integrity with an ability to create all required statistics for reporting with much less effort, than current paper-based version requires. The labor book is a document containing information about a woman in labor (e.g. age, residence, gravidity, parity, etc.), during labor (e.g. diagnosis, indication for surgery, surgery), the outcome of labor (Apgar score, pH) and immediate information about the newborn’s measures (e.g. length, weight) together with reported findings from neonatology department (traumatism, intubation, placement on ICU). Out of this data monthly statistics can be computed in selected areas of interest (e.g. number of surgical procedures, number of certain types of diagnosis, number of cesarean sections, relative perinatal mortality, etc). Given that in the University Hospital Brno there are performed approximately 5000 labors per year, it takes to create these statistics from paper labor books a few tens hours of work per month. Design of application is based on above mentioned ontology. In summary, the labor book contains, in all three above described parts, 54 parameters.

5.3. Design of application

Design of the electronic version of labor book is based on paper documentation, which is conducted in pre-printed blocks with defined fields. The requirement was to have the electronic version as similar as possible in structure and form to the original paper version. Unfortunately, in fact the records contained not only information that was prescribed for each form field, but also a lot of additional data. After gradual composition of materials and a series of consultations, we came to the decision to divide the area of interest into three basic parts: information about a woman in labor, information about the labor itself and information about a newborn (including information from neonatology).

The proposal is divided into two parts. In the first part we deal with access of single participants to the three basic parts of the system, as
5. The First Case Study: Obstetrics Area

mentioned above. This part of design is schematically illustrated using use case diagram. We use the diagrams for description of system, not for description of users access to application. In Figure 12 there is shown the basic use case diagram of links of healthcare personnel and a mother and a newborn. In principle, this process is divided into the administrative part of the admission of mother to the maternal hospital, and clinical follow-up of delivery and acceptance of a newborn to neonatology. Here we come to the first problem, which we will discuss in more detail in the following paragraphs.

As it is evident from Figure 12, there are four specialized medical professions in the whole process, namely obstetrician, neonatologist, midwife and nurse. Each profession has own role and the role defines validity of contained data.

![Use case diagram of process admission to maternity hospital, labour and admission to neonatology](image)

Admission process of the mother is schematically shown on the use case diagram of Figure 13. Firstly, medical staff is authorized and subsequently person to be admitted is identified and medical staff fills basic data about the person. This part (information about a woman in labor) is divided into
two blocks, namely general information (e. g. names, addresses and personal identification number – it is identification of the person) and information from admission (e. g. date and time of admission, gravidity count, parity). From the time perspective this block is filled with information as first. In terms of access to this part of record, it is entered by nurse, who performs admission of women in labor.

The next part is the description of labor itself. This part includes date, time of birth, week (+ days) of the pregnancy, in which the birth was, labor times, type of mechanism of delivery (spontaneous, caesarean section, etc.), information about anesthesia (no, general, epidural, etc.) connected to the diagnosis of birth, or indication for surgery and list of surgeries performed (general interventions). This part includes also information, which obstetrician led the labor, about midwife and physicians that performed surgeries. From time perspective this part is second in order. The information to this part of record is filled by midwife or obstetrician, who led the labor. This order is shown in Figure 14.

The third part contains information about newborn(s). This part is further divided into two sub-parts: The first sub-part includes a series of data that informs about the status of newborn after delivery (Apgar score, the blood pH and base excess and data about sex of newborn, length and weight. Second sub-part is information from neonatology, which contains mainly information on the possible trauma, neonatal intubation (and its length), location of newborn to the NICU and length of NICU stay, list of diagnoses and information about dispensary. This part is schematically shown in use case diagram in Figure 15.
5. The First Case Study: Obstetrics Area

Figure 13 Use case diagram of admission to maternity hospital

Figure 14 Use case diagram of Labour process with relation to observation, diagnosis and intervention (medication and surgery)
This construct is important for information exchange. Therefore, we present use case diagram from Figure 14 in HL7 version 3 model shown in Figure 16. Model is quite simple and it defines relationships to derive messages from labor and delivery outcome, given by observations, diagnoses, interventions and patients and physicians participate in these acts.
5. The First Case Study: Obstetrics Area

Figure 16 HL7 RMIM for Labor Book design

50
Overall summary of significant classes of applications is shown in Figure 17. The central class of the whole model is Labor, to which the Admission and Newborn classes are linked. This arrangement is given from original documentation (paper labor book) and has been retained for its clarity and uniqueness. A detailed description of these classes is mentioned in the paragraph about data model of the database.

Two basic classes are given in terms of delivery outcome. The first class represents result of delivery, which contains standardized set of possible outcomes. The second class is protocol of pathological delivery. This class is invoked only when there is satisfied some (always at least one) of the conditions that indicates pathological delivery (e.g. too much blood loss, low pH, low weight newborn). This protocol provides a summary of these conditions, the relevant epicrisis and expression of the doctors with the resulting recommendations. Schematically, the relationship of the classes is shown on Figure 18.

---

Figure 17 Class diagram of the clinical main part of labor Book Application
5. The First Case Study: Obstetrics Area

Figure 18 Class diagram of Labour results part with relation to the Labor and Newborn classes

Figure 19 shows relationship of classes role, personnel and user. The reason for this structure is the fact that not every personnel must be user and not every user must have a role in the delivery process. The second important thing is the ability to assign different roles. This relation arises directly based on HL7 RIM model. It allows the physician in one case to act as a leader obstetrician and in another case as a surgeon.

Figure 19 Class diagram of relationship user-personnel-role

To export data from the system there has been designed interface through five basic classes - Mother, Admission, Labor, Newborn and Neonatology (see Figure 20). Export of statistics is provided internally in the application. Statistics are divided into two levels. Basic statistics are given to the frequency of occurrence of given items in the records. These
statistics are made monthly and are used internally in the department of obstetrics. For conditional and complex statistics there serves frequency filter interface. This interface allows searching and filtering the records themselves and combined statistics are created on the same principle (e.g., number of caesarean births occurs simultaneously with pregnancy is higher than 3). The second integration interface is export to message.

![Figure 20 Class diagram for export of statistic and medical information](image)

### 5.4. Implementation of application

The data model is based on assumptions that were defined in the previous section. The basic unifying element is the record about labor. This is established in relation 1:1 to the admission of woman in labor, to the mechanism of delivery and anesthesia (according to the consultation, each birth just one drive and anesthesia - time does not play a role). Tables of diagnoses and labor are bound by the relation M: N with a parameter that determines whether a given diagnosis is an indication for surgery (indication for surgery is always one for one labor). Diagnoses are further distinguished into categories along data standard. We use Czech national data standard DASTA (http://ciselniky.dasta.mzcr.cz/), however codebook of diagnoses in this standards is the same as WHO standards ICD-10 (ICD, 2012) (only names of diagnoses are in Czech language and there are added some parameters for national statistical institution). This method of division was chosen because of simplicity of the direction to the application interface (for
diagnosis there is only one interface), while maintaining information about
the type of event. The same source (DASTA) is used for all codebooks in
application. The table intervention is linked to table labor in M:N
relationship. The term "intervention" is chosen deliberately because not all
acts that potentially fall into this area are surgeries (e. g. a blood
transfusion). Last link M: N is between tables and labor and personnel who
participated in the labor. As mentioned above, these professionals had been
divided into several categories, namely head obstetrician, midwife, surgeon,
and neonatologist. These categories ascend in the data model as the role and
are linked with employee table over this role table to labor table. At the
same time each employee is assigned to his/her expertise (doctor, medic,
midwife, nurse, pupil, other). The model is implemented as a relational
database in the database system PostgreSQL 9. 0.

The original intention was to use the electronic form for displaying the
information as well as for entering new data and their editing. This
intention, however, in the course of development proved to be unrealistic. In
analysis of the original paper labor book it was found that the records
contain not only information that is prescribed for each form field, but also a
lot of additional data. Therefore we have decided not to enable editing the
parameters directly in the spreadsheet form. The application was divided to
two parts. The first is a basic table, which displays selected information
from all three parts and is used as overview (see Figure 21). The second part
is an editing form with separate tabs for editing and filling information (see
Figure 22). Entered parameters are grouped into parts according to the
design and editing form can be gradually browsed by timeline, which was
also part of the proposal. The application is implemented in Java.
5. The First Case Study: Obstetrics Area

Figure 21 Main window of labour book application

Figure 22 Example of application editor dialog windows
Due to linking the applications with relational databases the issue of subsequent production of statistics is considerably easier. Basic statistics, which were created from paper documents, are the frequency of different parameters (the instance of diagnoses, mechanic birth, operations, results of births, etc.) in individual months of the year and their total annual number. The number of so far set of parameters is 97. It is obvious that generating statistics using database is better than manual processing - from the time perspective, but also from perspective of parameter combinations.

5.5. Partogram and Cardiotachogram

The previous description about labor book application is concentrated only to information without a timeline (as determined by the requirements of contracting authority). However, two time-dependent data entities are recorded during delivery. The first is the partogram. Partogram is time evolution of the delivery. Cardiotachogram is recording of the fetal heart rate and uterine contractions. This record is often associated with temporal events like medication or different properties of the signal (event type deceleration or acceleration). For this type of information, the primary model of labor book is not designed. One of the main aims of this work is to define an architecture that would allow such an extension. We do not want only to add to the record a reference to the file, in which the signal is. The aim is propose description of the signal in a form that satisfies the conditions of semantic interoperability, i. e. we want to use such signal processing and to convert the information from signal into a form of codes and identifiers. In the case of labor book the basic class of the signal model (Record) is associated with classes Labour (a recording during labour) and Newborn (measurements in newborns). Record class has three subclasses - events (e. g. medication), signal (contains a reference to the signal and important parameters of signal) and a class Complex containing those parts of the signal that are important (characteristic).

The record is complete with this added information. The Figure 23 shows the basic design of the connection of the signal model. We will deal with this proposal in more details in Chapter 6.
5. The First Case Study: Obstetrics Area

Figure 23 Extended class diagram. Proposal of addition of classes Events, Record, Complex and Signal for storage of Partogram and Cardiotocogram
6. Extended Approach: Proposal of Multi-layer Data Model

In this chapter, we describe the proposed approach to the representation and storage of information using multi-layer model.

6.1. Events

Events are defined as observation, measurement or finding some facts about the focused subject. We can consider as an event, for example, statement of diagnosis, medication, intervention, measurement or observation. In the case of the measurement, the event may be measured signals and/or the results of laboratory test, which is carried out on the basis of samples. On the other hand the observation is based on subjective evaluation from a person (expert). An event can be connected with certain vocabulary (codebook, nomenclature, other ontology). Ontology concept of event structure is shown as ontology in Figure 24. We have used Protégé [9] for creating of our proposed ontology.

In terms of temporal properties there can be defined two types of events, as we mentioned above. The first type is defined by its beginning and end (e.g. medication type infusion) has the same time length, second type of event is given in one moment (e.g. medication type bolus). The second option for the event is its definition as parametric (requires entering value and definition of unit) or nonparametric. Flow chart showing the branching when adding events is shown in Figure 25.
6. Extended Approach: Proposal of Multi-layer Data Model

Figure 24 Events part of ontology

Figure 25 Flow chart of event option at creation
6. Extended Approach: Proposal of Multi-layer Data Model

The concept of the events in our proposed model can be transferred to other areas. In Figure 26 there is shown the distribution of events in our model and in the model of HL7 standard.

![Figure 26](image)

**Figure 26** a) The dividing of events in proposed experimental model, b) related dividing in HL7

6.2. Signals, Points and Complexes

Signals are the foundation of modern medical diagnostics. With the discovery of ECG biomedical signal provides a basis of diagnostics. A set of invasive and noninvasive tests is very wide and number of signals that can be obtained during the examination can be large. The standards do not include form for decomposition of signal (in general) into semantic description.

In terms of standards for communication and information exchange the most common approach is to attach signals by reference. There are a number of formats for storing signals and a many of them have a complex form header (e.g. format MFER contains complete information about the type of signal, the sampling frequency, sensitivity and size of data block in header). This information is essential for subsequent processing, but it does
not allow semantic transmission. The problem is the variability of signals and the fact that more complex description in the header means more complex parser for the information retrieval. The signal contains a number of characteristic points and intervals, which may be important for subsequent evaluation of the signal. The identification of these intervals is the goal of most analyses. However, these intervals do not have to be given only by the nature of the signal. These intervals are typically associated with the physiological processes of the body. Events that affect the subject are addition to the characteristic interval. Of course these two definitions of intervals of the signal may overlap. For example, beginning and end of ventricular fibrillation can be determined directly from the signal. However, it can also be defined by saving events of beginning and end of ventricular fibrillation in the independent system with precise time synchronization. Thus defined points and complexes can be then represented using a semantic interoperability tools. In the next section we describe the design process, which defines storage of the signal with all the information and the transformation that converts the signal into a structured form. The basic concept of lead (signal) is shown in Figure 27.

![Figure 27 Lead definition](image)
Firstly, we define the term of lead. We suppose that lead (signal) can be still referenced or one signal or signal combination can be used as reference and data model considers more signal types. Therefore, protocol for specification of leads (signal-reference terminal) is defined, thus forming unified lead list from various sources.

Type of signal is a very important item of concept for following processing, because the next conditions of processing are dependent on type of the signal. This issue is discussed in the paragraph on feature extraction and evaluation.

Parameters of the signal contain basic information about the signal (e.g. sampling rate, sensitivity, maximum value, minimum value, file format). This information serves for a simple transfer of raw data and their reproducibility.

Time parameters determine the fundamental temporal properties of signal. The beginning and end of the signal are not the only important parameters. In many cases, the signals are recorded over a longer period of time with many important actions. Definition of time and value axes is very important for data integration (a failure of signal recording can happen and architecture has to allow for this possibility). We evidently find a representation of each signal at the appropriate time and in relation to other signals, parameters and events. All signals recording, segmentation and analysis are addressed through a general interface that allows to add new formats and methods of segmentation and analysis.

Once we have defined the term of lead, the next step is a definition of the basic part of the proposed model. This section defines concept of characteristic points, complexes, analysis and features for our model (see Figure 28).
We assume need of certain signal preprocessing. This step is also solved by means of interface in order to preserve defined variation of more signal sources.

The first way of defining Characteristic points is based on the signal. These points are the basic building blocks for the following description of the signal. As a typical example of characteristic points we can consider the characteristic points of the ECG beat (P, Q, R, S, T).

The second possible way of determining the characteristic points is an event. Events are recorded separately. Their projection on the signal is given by the time of occurrence. Events can be immediate or define certain interval. Hence, events are generally related to characteristic point class and also to complex class. Points (but points of event) define a complex and complex is defined by an event. The third link is given by the features that express a specific property signal and/or complex, respectively.

Complexes are defined on the basis of characteristic points. The basic form of complex is interval that is defined by two points (beginning and end). In general definition of complex we can use any number of points and their definition can be determined by their exact order, or just their presence in the complex.

Processing / Analysis are transformation and analyses intended to infer new signals (spectrogram, wavelet) and to an automatic determination of characteristic points for definition of complexes. Block of analyses is tied
back with a block of derived signals and with block of complexes. It means that processing / analysis can be used repeatedly for derived signals and signal complexes. These relations are schematically shown in Figure 29.

![Figure 29 Relation of Events with the Characteristic points and Complex definition](image)

Figure 31 shows part of the ontology dealing with signals, points and complexes, which summarizes all definitions mentioned above. In the following, we will describe some specific relationships in the signal part of the ontology in more detail. Firstly, we state definition of term *Lead* in the ontology. Class *Lead* is not directly present in the ontology. As shown in Figure 27, the *Lead* can be created from a measured signal and a reference signal (which is also usually defined on the basis of the measured signal). However, it is important to realize that many devices already store the measured signal in the form of *Lead* and reference is not always available. For this reason we do not define *Lead* directly (because it can be signal or derived signal), but we use for its definition the class *TypeOfSignal*. This class has a very important function for the ontology. In this case, it is this class that defines the signal (either directly measured or derived) as the
6. Extended Approach: Proposal of Multi-layer Data Model

**Lead.** The definition is shown in the following Listing 1 from the owl file. Example of inferred class hierarchy is shown in Figure 30.

**Listing 1 Measured signal class and ECG and Lead I in TypeOfSignal class**

```xml
<owl:Class rdf:about="&Ontology1356031541821;MeasuredSignal">
  <rdfs:subClassOf rdf:resource="&Ontology1356031541821;Signal"/>
  <owl:Restriction>
    <owl:onProperty rdf:resource="&Ontology1356031541821;hasType"/>
    <owl:someValuesFrom rdf:resource="&Ontology1356031541821;Lead_I"/>
  </owl:Restriction>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:restrictions/>
      <owl:equivalentClass>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <owl:Class rdf:about="&Ontology1356031541821;ECG"/>
            <owl:Description rdf:about="&Ontology1356031541821;Lead_I"/>
          </owl:unionOf>
          <owl:someValuesFrom>
            <owl:TypeOfSignal/>
            <owl:equivalentClass>
              <rdfs:subClassOf rdf:resource="&Ontology1356031541821;TypeOfSignal"/>
            </owl:equivalentClass>
          </owl:someValuesFrom>
        </owl:Class>
      </owl:equivalentClass>
    </owl:Class>
  </owl:Class>
</owl:Class>
```

![Figure 30 Class hierarchy and inferred class hierarchy for Lead of signal](image-url)
6. Extended Approach: Proposal of Multi-layer Data Model

As shown in Figure 28, the signal is the basis of this part of ontology, which comes under the event of measurement. This is due to the time range of the signal. The signal is in digital form (has sampling and quantization parameters) thus consists of clearly defined points (SignalPoints). Some of these points may be identified as important for the description of the signal and then they are defined as characteristic points of signal (CharacteristicSignalPoints). Identification of these points can be based on expert knowledge (are marked in the signal) or based on automatic/automated segmentation of signal and - last but not least - characteristic points are connected with events (e.g. diagnosis). This last relationship is reciprocal, because points can define (e.g. based on segmentation) the event. We present brief theoretical example: At measurement of EEG (evoked potentials) there is used (visual) stimulation. This stimulation is defined as an event type of intervention and it is marked by the user at the time of stimulation. In the measured signal there are relevant responses detected (e.g. P100 or P300). Another example is spontaneous activity in the form of an epileptic seizure. This event was not stimulated by the user, but it was found at a processing of measured signals. The found event is identified and subsequently labeled as event type of diagnosis - epileptic seizure.

The definition of intervals and complexes using characteristic points has already been mentioned above. It is important to mention one more feature that has the complex - can contain other complexes. Definition of Interval and Complex in OWL is shown in Listing 2 and Listing 3.

Listing 2 OWL class interval

```xml
<owl:Class rdf:about="#Ontology1356031541821;Interval">
   <rdfs:subClassOf rdf:resource="#Ontology1356031541821;Complex"/>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#Ontology1356031541821;hasCharacteristicPoint"/>
         <owl:onClass rdf:resource="#Ontology1356031541821;CharacteristicSignalPoint"/>
         <owl:qualifiedCardinality rdf:datatype="xsd:nonNegativeInteger">2</owl:qualifiedCardinality>
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
```
6. Extended Approach: Proposal of Multi-layer Data Model

Listing 3 OWL class complex

```
<owl:Class rdf:about="#Ontology1356031541821;Complex">  
  <rdf:subClassOf rdf:resource="#Ontology1356031541821;SignalDescription"/>  
  <owl:Restriction>  
    <owl:onProperty rdf:resource="#Ontology1356031541821;hasCharacteristicPoint"/>  
    <owl:onClass rdf:resource="#Ontology1356031541821;CharacteristicSignalPoint"/>  
    <owl:minQualifiedCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">2</owl:minQualifiedCardinality>  
  </owl:Restriction>  
</owl:Class>
```

Figure 31 Signal part of ontology

The ontology in Figure 31 illustrates the fact that complex can contains complex by relation loop (relation hasComplex) in Complex Class (same construct is also used for class ProcessedSignal, because the derived signal can be input for another derived signal). In the ontology there is the difference between interval and complex given by the number of points that must and can contain. From the constructional point of view, the interval has to contain exactly two characteristic points and the complex has to contain at least three characteristic points.
6. Extended Approach: Proposal of Multi-layer Data Model

The definition of intervals and complexes using characteristic points has already been mentioned above. It is important to mention one more feature that has the complex - can contain other complexes. The ontology in Figure 31 illustrates this fact by relation loop (relation hasComplex) in Complex Class (same construct is also used for class ProcessedSignal, because the derived signal can be input for another derived signal). In the ontology there is the difference between interval and complex given by the number of points that must and can contain. From the constructional point of view, the interval has to contain exactly two characteristic points and the complex has to contain at least three characteristic points.

This part of ontology is important and we can verify this with other concepts. We use mapping to HL7 aECG conceptual model [81] and ECG reference ontology [80]. Figure shows mapping to aECG conceptual model, figure shows mapping to ECG ontology model on the side of the physician. In [80] is presented also mapping to AHA/MIT-BIH and SCP-ECG conceptual model. Therefore, we suppose validity of our model with these two models too. The mapping with aECG is shown in Figure 32, the mapping of our ontology and ECG reference ontology is shown in Figure 33.
6. Extended Approach: Proposal of Multi-layer Data Model

Figure 32 Mapping between the HL7 aECG conceptual model [81] and our model

Figure 33 Model of the ECG on the side of the physition of ECG reference ontology [80] and signal part of our proposed ontology
6.3. Evaluation and description of signal and data

An evaluation of the record or part of the record is the diagnosis relevant to the given part of record. This event can be seen as an expert evaluation because decision is (in our concept) made by physician. We want to give a description of a signal, which corresponds to the given sections of this evaluation. The basis of this description is processing and analysis of the resulting signal. Results of this process is segmentation of signal, transformations of signal between domains (e.g. from time to frequency) or transformations to other signals (e.g. the independent component analysis (ICA) - transformation into source signals). The result of transform is defined as ProcessedSignal (in ontology) and it is again signal (loop closure occurs).

Artificial intelligence uses for signal description the term "feature" (and the associated area pattern recognition). Using methods of pattern recognition a classification task can be performed (over a set of features), i.e. task of assignment of an instance to one of the defined classes. But, in our case, the result is not only at the level of classification of diagnoses identified by expert (imagine this process in hierarchy on upper level). In part associated with the signals and their segments and analyses we want to use features to description of the individual parts of the signal (imagine this process in hierarchy on lower level). The aim is to describe signal at the semantic (aggregated) level for each particular portion of the signal (as points, and complexes). Subsequent classification at diagnosis level should be given by combination of descriptions of individual parts of the signal (with appropriate conditions). The resulting classification can be achieved directly, but our purpose is to create an intermediate layer that allows to interpret the processed parts of signal (described in the form of features) in relation to evaluation (there may be rules) as well as in relation to the final diagnoses (again there may be rules). In other words, if the model is defined to represent the signal and relationships between events and signal, the next step is the transfer of this information to a semantic form. And here we present the simplest form of the description of a signal using rules which are often used in the medical field for simple interpretable.. The rules represent background knowledge of given area. The basic layout of the descriptive
part illustrated in Figure 34. The system is composed of four parts – Premise (antecedent), Conclusion (consequent), Rules, Description and Evaluation:

- Premise – set of condition in standard form like complex_length > value_A, complex_maximum < value_B, point_value = value_C and so on. Condition is related to signal type. i.e. for specific signal type exist specific set of condition;
- Conclusion – statement of fact about a given point, complex or signal; it is list of evaluation options; categorical classification of point, complex or signal;
- Rules – condition or conditions and results in standard form: if condition then result;
- Description – assignment of evaluation type to a given item of record (rules for a signal);
- Evaluation – represents a step away from the lower layers of knowledge, thus from the description of the signal, to knowledge on the level of diagnosis; it has its own rules, which are designed to combine the results of the rules of the bottom layer.

Example of rules for signal complex description is shown in section 7.5, where there is a simple example of description of electrocardiogram (EGC) complexes and points.

Signal description is transformation of information to knowledge. This transformation represent the intermediate layer of knowledge that in created based on data because the signal description is connected with rules evaluation. Our knowledge is not only in a one layer, but in two layers - diagnosis and description of measurement. It is obvious that there should be defined a possibility of working with uncertainty as in the top layer (diagnosis) as well as in the bottom layer (description) of knowledge. Form of uncertainty is limited by our knowledge representation (in our case it is the uncertainty when working with rules, such as weights, measures, trust, and distrust). The basic model of signal evaluation is shown in Figure 34.
The given rules are formed for each signal type separately. Sets of rules are therefore dependent on whether the processed signal is electrocardiogram (ECG), intracardiac signal, continuous recording of blood pressure, electroencephalogram (EEG), etc. Rules should be hence categorized according to the usability and information should be maintained, in which signal domain rules can be used. The rules for the description of ECG and description of EEG will be different, not only in parameters, but also in the domain, in which the description is carried out (for ECG rather the time domain, for EEG rather the frequency domain). Part of the description is not only information about the values of features, but also about the domain, from which the feature came, which ontological structure allows. Scheme of signal types, signals, complexes and points, rules and evaluation is shown in Figure 35.
6.4. Multi-layer model

We have described parts of the multi-layer model. Now we describe their connection to the resulting model of representation and data storage.

The class diagram of complete model is shown in Figure 36. Part that defines leads and signals is called Lead layer. Part of preprocessing, processing, analysis, and complexes of characteristic points is called Analyze layer. Part that contains conditions, rules and evaluation is called Evaluation layer. Block of events is called Event layer. Lead layer and Analyze layer are connected simply via leads definition and their preprocessing, which is already defined in the Analyze layer. Furthermore, type of signals in Lead layer is connected to conditions in Evaluation layer (see also Figure 37). Definition of leads is linked to the block Evaluation in
Evaluation layer, which allows direct evaluation of the actual signal. The last link from Definition layer is to the Event layer. This relationship defines the ability to assign an event to the whole signal. Analyze layer is connected to the Evaluation layer over blocks Characteristic points and Complex definition. This relationship is fundamental to the evaluation of the signal. Complexes are also connected to Events as an extension of the event itself. Evaluation is further connected to the block of derived signals.

Complete ontology model is shown in Figure 37, where all above mentioned parts are linked into a whole. This model is only an abstract (not assigned to any specific process or environment) and so all activities are derived from the existence of the subject (basic entity). Main class (within the meaning of documentation) is the class Record.
6. Extended Approach: Proposal of Multi-layer Data Model

Figure 36 Class diagram of Multi-layer model
Figure 37 Ontology of Multi-layer model
7. The Second Case Study:  
Collection of Data at Medical Experiments

In Chapter 5, we tried to approach design applications of clinical practice. We proposed architecture extending the current data model by the record of partogram and cardiotocogram. In this chapter we will describe this approach on the application to a medical experiment issue. This section contains not only one application, but it is divided into three applications. Two of them are part of our proposal and we will describe them in more detail. In this case study we use architecture proposed in Chapter 5 and present the possibility of its use.

7.1. Multi-Layer Model of Experiment

The experiment ontology is based on ontology of multilayer model (from previous chapter), but the main class is Experiment. The idea of this step is following. If we describe the abstract model we use as the start class a record and a subject and the record and the subject is not placed into any environment (the model is abstract). However, if we have a specific process in a given environment, all the descriptions are linked to this process. Therefore the primary classes are Experiment, and Subject (in our case class Swine). In other words, if we have not specific process in the given environment we can describe only the subject and record but if we have a specific process we have to define all classes in relation to this process.
Important subclass, which is added to the experimental ontology is *Scenario*. It is a subclass specific for the experiment. It is clear that each experiment must have scenario and we divide the structure of scenario into branches and phases. Example of subject and person in ontology is shown in Figure 38.

### 7.2. Assumptions of the system

Before we describe design and implementation of applications used in the experiment, we summarize some basic facts and requirements derived from that domain. We will talk about the specific environment with specific requirements and options, which again define the constraints of our implementation.

Experimental laboratory environment is based on several permanent devices, which form the basis of information about the observed entity. In particular there are bedside monitors, devices that provide important physiological data (heart rate, pressure, blood oxygen saturation, etc.) as signal, trend of values and immediate value. Only signals and trend values are automatically recorded.

Another device is mechanical ventilation. Mechanical ventilation is necessary in our experiments and has number of regimes and recording of important parameters of ventilation.
Next important set of parameters are results of laboratory tests (e. g. blood analyzer used for basic analysis of acid-base balance).

Because the most experiments are linked with the electrophysiology field, devices for recording of intracardiac signal and intracardiac stimulation are used.

Moreover, special devices (important for experiment) can be added and it is necessary to add them also to the data store system. In this case, it is important whether the device is equipped for communication with external devices (the communication protocol, the protocol is determined by an internal or standard). Simple diagram of basic devices is shown in Figure 39.

![Figure 39 Schema of basic arrangement of measurements in The Experimental Laboratory of Electrophysiology](image)

7.3. Design and Implementation of Applications

Firstly, we state an overview of events that take place in the beginning of the experiment. The main participant in the initial processes of experiment is a doctor who provides basic tasks. Veterinarian is in the role of advisor mainly in the field of medication and for autopsy. Engineer or
technician is responsible for setting up the devices and applications for recording and data collection. These relationships are shown in Figure 40.

![Figure 40 Basic use case of initial processes of experiment](image)

The second model, which we state in the introduction to this work is presented in an abstract level use case diagram (see in Figure 41), showing each participant access to events during the experiment. Physician specialist is added to the previous participants, who has a major share in the course of the experiment, and basically he/she controls the experiment.

![Figure 41 Use case diagram of experiment](image)
7. The Second Case Study: Collection of Data at Medical Experiments

These measurements represent complex electrophysiology investigation particularly useful in the case of experimental measurement. Along with the data about the state of a measured subject (age, anamnesis, medication, etc), we obtain very large database for complex analysis of heart function. This paper presents solution of the first step of heterogeneous data processing – design of basic data model and implementation of an application (on the designed data model) for integration, pre-processing, analysis and visualization of data. Data in the data model can be consequently used for analysis of functional relations of acquired data. This project has been solved by an interdisciplinary team of experts in biomedical engineering and medicine. Ontology part of roles is shown in Figure 42.

![Personnel part of ontology](image)

**Figure 42 Personnel part of ontology**

7.4. Events

Automatic recording of measured data is very important for a following assessment of the experiments. However, during the experiment there happen many events. A large part of events cannot be recorded automatically. Therefore, there is needed a tool that enables easy route registration of these events. Before we get to the actual method used to record events, we describe the concept of an event and division of events. Example of events of experiment in the ontology is shown in Figure 43.
Generally, it is defined by hasCause relationship of the Event class. In our case, this relationship is used several times as an example of five relations in cardiac arrest shown in Figure 43. The first relation is defined between electrophysiology stimulation and ventricular fibrillation because stimulation is the cause of ventricular fibrillation. Next relation is between cardiac arrest and ventricular fibrillation because ventricular fibrillation has consequence in cardiac arrest (in other word ventricular fibrillation is cause of cardiac arrest). Next two relations are between cardiac arrest and cardiopulmonary resuscitation and extracorporeal membrane oxygenation (which are interventions for sustaining life). And the last relation is between cardiac arrest and medication bolus of adrenalin, which helps to keep better function of organism in shock state as cardiac arrest.

The term event represents an act, which is defined by a process of living objects or observations on living objects. Events are divided into five groups: diagnoses, medications, interventions, measurements and other events. Events can be divided into events that have beginning (OPEN) and end (CLOSE), and events that have a start and end at essentially same time (OPEN AND CLOSE). Of course, some events can be identified in the data. However, their identification may be difficult and time-consuming. Moreover events can form a corpus of experiment, the scenario that
provides maximal consensus of single experiments (in the same area). Therefore, the framework of events is divided into two record types. The first type of event record is the scenario, where each event is unique at a given time after it occurs and is inaccessible for the next record. According to this scenario, it is then possible to browse the experiment always in the same sequence of events. Plain events are second type of events that can be repeated to an unlimited extent, and that represent unexpected events or events where it is not possible to pre-define a number of repetitions. Single instances of events may be linked to other events in the sense that their arrival automatically closes these defined, already running events. In our case, tool for recording events is created by an application, in which events are recorded via a computer with a touchscreen. Sample of application is shown in Figure 44. One of the most essential parts of the whole concept of integration is the data model of the issue. The model is divided according to individual events that relate to the subject experiment.

![Figure 44 Example of event record application](image-url)
7.5. Signals

Example of signal part of experiment in ontology is shown Figure 45. We present an illustrative example of a simple inferring for the assignment of the characteristic points to parts of a signal and thus to define these parts as complexes and intervals with their mutual relations. In the Figure 46 the hierarchy of subclasses in class Signal is shown. There is shown an example of signal Signal_1. In the class ProcessedSignal there are defined five structures that are created from Signal_1 by signal segmentation. In the right part of the Figure 46 is shown Inferred Class Hierarchy, which already shows the assignment of Signal_1 to Lead_1 and individual structures under Signal_1. By assigning characteristic points to structures of signal defined complexes and intervals are assigned. In Figure 47 a) there is shown Class Hierarchy of complexes and models. The Figure 47 b) shows then Inferred Class Hierarchy where the individual structures of Signal_1 are assigned to the corresponding complexes and intervals. The Figure 48 shows the relationship signal, its structures and complexes to characteristic points and intervals in the ontological graph.

Figure 45 Example of signal part of experiment in ontology
7. The Second Case Study: Collection of Data at Medical Experiments

Figure 46 Class Hierarchy and Inferred Class Hierarchy for signal and its structures

Figure 47 Class Hierarchy and Inferred Class Hierarchy for complexes, interval and signal's structures
7. The Second Case Study: Collection of Data at Medical Experiments

We suppose that signal can be still referenced or one signal or signal combination can be used as reference and data model considers more signal types. Therefore protocol for specification of leads (signal-reference terminal) is defined, thus forming unified lead list from various sources. The application has steps of signal loading and processing solved by means of the Interface (in accordance with Java definition) that allows use of various algorithms (example can be seen in Figure 49).

The part of signal preprocessing is very important is. For example, ECG signal is preprocessed according to recommendation and this recommendation is in preprocessing definition.

Before we start signal preprocessing we need to distinguish between signal and single data values (usually data exported from a data source). It is expected that the information content in the signal is relevant to the whole event and/or it is possible to divide a signal into several events and assign more signals to one event (it is important to say that length of the signal is often shorter than time interval between events). Within an event, there can be performed change of time resolution of the signal. It is possible to perform analysis that generates representative attribute(s) of signal information (descriptive feature/s). The goal of this definition is clear separation of data that have only a single value within one event or can be

Figure 48 Ontology graph of relations among signal, its structures, characteristic points and complexe and intervals
7. The Second Case Study: Collection of Data at Medical Experiments

replaced by one value without any loss of information. Schema of signal processing, storage and visualization is shown in Figure 50.

![Figure 49 Signal reader interface with time axis creation](image1)

![Figure 50 Scheme of signal processing, storage and visualization](image2)
Definition of time and value axes is very important for data integration. Signals can remain in their original format. Timeline is formed on the basis of information about the beginning of the recording, the length of the signal and the sampling frequency. By creating this timeline we can then control the analysis and visualization of data and events.

Signal viewer is basic part for visualization of measured signals with possibility of callipers application for measurement of intervals lengths and signals amplitudes, averaging and marking of characteristics points in the signal. The characteristic points are defined as function points (e.g. ECG points), as mark of some event in signal (e.g. ablation) or beginning and end of some interval (e.g. fibrillation). The example of signals with marking of ECG characteristic points is shown in Figure 51. Data model supposes composition of complexes (interval in the simplest case) from defined points. The way of composition is given as combination of points with number of order in the complex and definition of the complex. This simple way provides general approach for creation of complexes based on function aspects and/or user definition. Complexes together with points represent basic elements for analysis. The analysis means extraction of information from signal into descriptive features.

To the part of signal displaying, there belong also functions of advanced visualization of the signal and its properties that can be also used for analysis of the signal. These additional components of signal viewer are mainly trends of statistic parameters, overview of calculated signal parameters (correlation, coherence, etc.) and visualization of time and frequency signal properties.
The last part of application is analysis of signal and its parts. The analysis is basic process for description of signal and feature extraction. The analysis has the same assumptions as signal readers and pre-processing steps. Thus results of analysis are transferred to model using the transforming interface that allows use of larger number of various algorithms. It is important that the application is oriented mainly to integration of data (it does not need to include all analysis algorithms). Hence there is defined a possibility to transfer analysis results from external source (with storage of analysis type and its parameters). The results of analysis and information about long-time and middle-term values are used in evaluations, which are in many cases well-founded by rules from previous analysis or knowledge of experts. This process is not one-stage and we suppose that the evaluation table will contain several level of evaluation. Very simple example can be the following one:
7. The Second Case Study: Collection of Data at Medical Experiments

- simple level of evaluation is “QRS complex is normal”
- in rules it means the length of QRS > 0. 0. 6s and < 0. 1 s;
- higher level of evaluation is “signal is normal”,
- in the rule from previous evaluations it means QRS complex is normal and QTc interval is normal, in basic rules it means length of QRS > 0. 0. 6s and < 0. 1 s and length of QTc ≤ 0. 40s.

Window with conditions, rules and evaluation is shown in Figure 52.

![Figure 52 Example of conditions, rules and evaluation creation](image)

The basis of all code values are code books that form the basis for encoding data on the state of the subject, but also for the definition of evaluation signals (characteristic points, points for obtaining interval, the values of amplitude, marking events) and last but not least, for the definition of laboratory tests carried out, measured parameters and units.

Knowledge Discovery and Data Mining (KDD) methods require extensive data bases as an input. In cardiology we can acquire large volumes of heterogeneous patient data using different time scales. This is the basic reason that has led us to design of a data model, database architecture and processing steps that would allow efficient and transparent collection of heterogeneous data with different time scales in one storage and successive processing using different methods for different data types. It is process of
original data set transform into strictly defined form.

7.6. Class Diagram of Multi-layer Model

The main result of this application is integration of basic patient data (age, gender, weigh, high, systolic and diastolic pressures), information about diagnosis, anamnesis and medication, laboratory results and measured signals. Further we define process for signals segmentation in order to acquire time structure as interval and complex. From both these elements (generally points, complexes) we can perform analysis for extraction of signal description useful for evaluation of signal part and signal as complex respectively. Thus we acquire comprehensive overview of the measured subject and we can extract related information for more complex evaluation. A data model of the given problem is a very important part of the whole concept. The model has been derived from the following assumptions that were formulated based on consultations with medical doctors from the area of physiology and clinical cardiology. Basic unifying element of the record is the subject (patient). Basic entity having time specification is an event. It must be possible to attach to each event any number of signals (with respect to signal type, sampling rate, sensitivity, processing). A component of laboratory test results (type of tests, measured values and units) is formed analogously to results of each event. Information about long-term (e. g. height, diagnosis, personal and family anamneses, medication) and middle-term (e. g. weight, systolic and diastolic pressures, medication) state of subject is related to the measured subject and not to each event. All code values are defined in code books – in addition to above mentioned values there can be included also values for signal segmentation, analysis and evaluation - characteristic points (generally marks of any event in signal), points for definition of complexes (the simplest complex is interval), analysis value (amplitude value, area value, etc.). The model is shown in Figure 53.
7. The Second Case Study: Collection of Data at Medical Experiments

Figure 53 Multi-layer model for data from experiments in The Experimental Laboratory of Physiology
8. Main Contribution and Achievements

We have presented a proposed general ontology and data model, which allow semantic description of a signal. This is the main achievement of the thesis. It is necessary to stress that the ontology and model can be used for any type of measured signal independent on its morphology, temporal and frequency characteristics and the signal (or its parts) can be linked to events occurring during the signal measurement. The signal can be structured by segmentation to parts and individual points, which allows its representation in symbolic form with direct link to events, which occurred during signal measurement. Symbolic representation allows to create a link between machine processing and knowledge of expert, because information is defined in context of both signal and related events and moreover it is readable and understandable both for a computer and a human. The ontology model is defined generally for any biological signal. We have validated the proposed model for clinical parameters followed in clinical practice by mapping the proposed ontology to already verified models of HL7 aECG concept (see Figure 32) and ECG reference ontology (see Figure 33) and by definition of a specific instance of the model for representation of medical experiment structure. Symbolic representation of signal with connection to clinical events is useful in particular in the area of medical experiments.

In chapters 3 and 4 there is presented research overview from the area of standards in healthcare (Chapter 3) and existing methods for heterogeneous
8. Main Contribution and Achievements

data processing (Chapter 4). They indicate that nowadays there does not exist any model, which allows semantic representation of general biological signal and its direct connection with clinical values. There are models, which describe only a specific type of signal, namely ECG, on the level of semantic interoperability. However, these models are not transferable to other type of signal. Their transferability is blocked by strict definition of the signal structure, as ECG components P wave, QRS complex, T wave and so on. Thus other more detailed descriptive elements as points, intervals, shapes, etc. cannot be expressed by these models. In addition, they do not have the option to link the signal with events. These facts motivated us to propose a more general model that would be applicable to any type of signal.

In chapter 5 we proposed a model (ontology), which is based on a case study describing the clinical setting – in gynecologic-obstetrics department – that defines representation and structure of clinical events. In this case we performed an analysis of obstetrics area and defined a domain model (ontology). We validated the model in the application of electronic labor book, which has been designed for practical application, and is deployed at the University Hospital, Brno. At the end of chapter 5 we have discussed the issue of inclusion of signals into the presented clinical information system.

Based on considerations discussed in chapter 5 we have defined an extended version of the model, which allows semantic description of a signal (type is not important) and its direct connection with clinical events. This new concept has been described in detail in chapter 6, showing not only static description of the ontology structure, but also processing of the signal and its relation to the structural description. There has been performed mapping of our proposed ontology to the concept of HL7 aECG model and ECG ontology, which has proved the possibility of using our proposed model to representation of specific type of data. Moreover, we have extended the model by a new layer that represents a middle layer between expert knowledge and signal description for computer processing. The content are rules assigned to symbolic signal description. The model (ontology) generalizes the structure by dividing into three parts (characteristicPoints = important points in the signal, interval, and complex), which allow to reach the same results as a firmly defined
structure (this process is described in chapter 6 and schematically shown in Figure 33). This construct also allows mutual interaction with clinical events and serves as a basis for feature representation as defined in artificial intelligence.

In chapter 7 we have used the proposed model for representation of structure of participants, events, measured parameters and signals at a specific medical experiment in the area of extracorporeal circulation and electrophysiology. This case study has proved the importance of more general model of the signal and related events because these experiments generate large volumes of interrelated data and signals that have to be adequately represented for future evaluation of the experiments. One of the most important issues is preservation of the mutual links between signals and events.
9. Conclusion

The main aim of the thesis was to design an architecture, in accordance with the requirements of procedural and semantic interoperability in healthcare, in signal field linked events. This architecture is linked to existing solutions in the fields of nomenclatures and standards in healthcare. Its main part is the multi-layer model that contains part of events (acts), part of signals and part of evaluation that representing background knowledge. To achieve this aim, we divided it into several particulars aims in Section 1.2.

1. propose ontology model for heterogeneous data (clinical information and signal) with emphasis on the representation of the signal.

The proposed ontology is formal representation of our model. We define the knowledge about single part of documentation with request to semantic interoperability. The main part of our work in ontology is design of relation between event and signal part and evaluation of the signal part (interval, complex). We compare our proposed ontology with HL7 aECG conceptual model and ECG reference ontology.

2. propose multi-layer data model for heterogeneous data containing common clinical data, events and signals together; design the signal part of model with support of semantic interoperability
9. Conclusion

The core of the thesis is design of multi-layer model that is described in Chapter 6. We have specified the main distribution of data types and define the structure for their representing. Thus, we have obtained architecture, which allows defining general concept of signal leads with all necessary information, decomposition of this signal into important complexes using of defining of characteristic points and derivation of new signals. This signal domain is connected to the background knowledge, representing by compilation of signal type addicted rules. Our model defines two levels of knowledge which allow assessment of signal itself and also parts of the signal and overall evaluation of the state. And our model allows also linking to other information. Moreover, it is important to realize how documentation is valuable in the case that it is taken into consideration that data can be extracted also retrospectively. For such a procedure it is often necessary to have not only aggregated data but also the original data (signal).

3. in the obstetrics area specify the important part of domain model for description clinical process - main areas of clinical information – in our model

We have developed a system in Obstetrics area that allows to record clinical data from labour (viz. Chapter 5). At this process we specify the important part of Obstetrics domain model, thereby we have tested model of basic areas of clinical information. For the classic description - in terms of electronic medical records - it is very important to define its content in structured form. In Chapter 5, we present led such a proposal, and then we ask led a question how to extend this record on the signals (see next paragraph). At the end of the Chapter 5, we examined the possibility of adding information from measured signals during delivery. We have arrived at the simple concept of adding the package classes that represent such information. It is clear that simply attaching the data signal with a reference is still more unsatisfactory, in terms of increasing requirements on stored data. It is necessary to define a model that
allows representation of large quantities of data (which signals represent) in a clear description, relative to the context of the entire documentation.

4. define and develop this model in specific area of medical experiments

We have defined and implemented the multi-layer model in The Experimental laboratory of Electrophysiology. Applications (described in Chapter 7) were implemented for recording of designed events, integrating of measured signals and parameters, their visualization, interfaces for processing and analyzing. Proposed model is at least as important for the medical experiment as for clinical practice, perhaps more. Experimental environment is the best opportunity for testing the model, because it supposes using the whole spectrum of measured and stored data. Therefore our conclusions are demonstrated on medical experiments.
Bibliography


[43] Continua Health Alliance, "Continua Overview Presentation, The Next


[47] A. Levy, A. Rajaraman and J. Ordille, "Querying Heterogeneous Information
Sources Using Source Descriptions," in Proceedings of 22th International
Conference on Very Large Data Bases, Mumbai, 1996.

[48] M. Gagnon, "Ontology-Based Integration of Data Sources," in 10th
International Conference on Information Fusion, Quebec, 2007.

and S. Hübner, "Ontology-based Integration of Information – A Survey of
Existing Approaches 108-117," Proceedings of IJCAI-01 Workshop: Ontologies

Classification With Dependence Trees," IEEE Transaction on Geoscience and

[51] C. Chirathamjaree, "A Data Model for Heterogeneous Data Sources," in IEEE

[52] V. Kashyap and A. Sheth, "Semantic Heterogeneity in Global Information
Systems: The Role of Metadata, Context and Ontologies," ACM SIGMOD
Record, vol. 28, no. 1, pp. 5-12, 1996.

[53] A. Katifori, C. Halatsis, G. Lepouras, C. Vassilakis and E. Giannopoulou,

[54] N. Kannathal N., U. Acharya, E. Ng, S. Krishnan, L. Min and S. Laxminarayan,
"Cardiac health diagnosis using data fusion of cardiovascular and
haemodynamic signals. Computer methods and programs in biomedicine,
Bibliography


