Semantic Description of Health Record Data for Procedural Interoperability

Jan Vejvalka[1], Petr Lesný[1], Tomáš Holeček[3], Kryštof Slabý[1], Hana Krásničanová[1], Adéla Jarolímková[2], and Helena Bouzková[4]

[1] Faculty Hospital Motol and 2nd Faculty of Medicine, Charles University, Prague, Czech Republic
[2] CESNET, z.s.p.o., Prague, Czech Republic
[3] Faculty of Humanity Studies, Charles University, Prague, Czech Republic
[4] National Medical Library, Prague, Czech Republic
{jan.vejvalka,petr.lesny,krystof.slaby}@lfmotol.cuni.cz

Abstract. Growing volume of knowledge that needs to be processed and communicated leads to penetration of information and communication technologies (ICT) into biomedicine. Specialized tools for both algorithmic processing and for transport of biomedical data are developed. Proper use of ICT requires a proper computer representation of these data and algorithms. We have analyzed the openEHR archetypes in order to utilize openEHR formatted data in medical grid environments on the MediGrid platform. Both openEHR and MediGrid utilize the phenomenological approach to biomedical data; however the level of constraint placed by both systems on the concepts transported or processed data is different.

Keywords: openEHR, archetypes, MediGrid, phenomenology, medical algorithms, health records, semantics, ontologies.

1 Introduction

Growing volume of scientific knowledge and resulting specialization in medicine stresses the importance of setting and implementing algorithms and guidelines, which objectify and standardize healthcare processes. Despite the differences in understanding of standardization and algorithmization between the world of medical practice and the world of information and communication technologies (ICT), ICT are often used to support standardization in medicine. Proper use of ICT requires correct computer representation of relevant data and algorithms. Several successful projects aiming at creating or documenting a collection of biomedical algorithms exist (such as MEDAL [5], MedCalc etc.).

OpenEHR [1] is Open Source Electronic Health Record architecture, based on the strict ontology description principles. OpenEHR is based upon archetypes – computable expressions of a domain content model in the form of structured constraint statements, based on reference (information) models of clinical concepts.

P. Kostkova (Ed.): eHealth 2009, LNICST 27, pp. 124–130, 2010.

MediGrid [2] is an open-source algorithm description system which was developed in reaction to clinical requirements for correct biomedical data algorithm processing. It minimizes the possibility of errors caused by most common misuses of clinical data in biomedical algorithms (cf. Table 1) by providing tool to manipulate the domain ontologies of biomedical algorithms.

Table 1. Common misuses of biomedical algorithms on clinical data

Error	Examples
Application of good algorithm to wrong data	*Trivial:* entering weight in pounds into the common BMI formula. *Less trivial:* using incorrect formula for body surface area (more than 4 formulae exist, each giving different and in the case of small children even wrong results) for drug dosage.
Selection of wrong algorithm	*Trivial:* using standard deviation instead of percentiles in normalization of body weight *Less trivial:* use of body mass index percentile instead of the weight for height ratio percentile for proportionality in early post-pubertal girls, leading to false statements of obesity.

The purpose of our work presented in this paper was to postulate the principles of MediGRID as a system for semantic-based representation of biomedical knowledge and to compare MediGRID and openEHR. We compared them from the perspective of practical applicability in knowledge domains corresponding to specific areas of medicine, aiming to find possible complementarities in the respective approaches.

2 Materials and Methods

MediGrid is being designed to facilitate analysis of semantics automatic processing of semantically well described biomedical data with scientific knowledge represented by semantically well described biomedical computational algorithms, with the means of up-to-date information technologies.

MediGrid approach aims to representation of biomedical data and algorithms as resources, and on sharing these resources in a grid-like environment. The natural way of assigning biomedical algorithms to data (or vice versa) is by comparing their semantic values; therefore MediGrid builds its mechanisms of sharing on capturing and processing of semantic information. Results from previous projects (SMARTIE, Growth 2) have been used as a base for our work: we used existing algorithms to verify our results and we used our experience from SMARTIE to structure semantic information.

The basic assumptions of MediGrid are:

- Data processed by biomedical algorithms are (following the philosophical tradition of phenomenology as formulated by Husserl [3]) indicators[1] that can be transformed into other indicators and grouped into indicator classes by their roles in these transformations.
- Data and algorithms can be shared across conceptual domains if trusted semantic links exist to support such interconnection.

We have performed retrospective analysis of current representations of algorithms and algorithm repositories and formulated several key principles, based on a sound philosophical background, on accepted principles of documentation and scientific communication, on which our approach is based. We have implemented these principles in software and validated their usefulness and effectiveness on the application domain of paediatric auxology.

3 Results

3.1 Description of Medigrid Entities

The need for extensive review and verification of semantic links when matching data with algorithms, and also the stress on correct procedures in medicine (lege artis) result in practical requirements on design and implementation of a knowledge-processing system based on these theoretical principles:

- Semantic information (meaning for the human user) of both indicator classes and transformations must be explicitly described and readily available for user assessment and validation.
- Semantic information must be bound to the current scientific paradigm and to evidence based medicine through extensive links to published and reviewed works.
- Mechanisms of procedural authority and trust must be implemented to support users' decisions about procedural values of individual components.

We have identified several logical layers, which need to be documented in order to meet the requirements mentioned above:

- Source layer, which contains the description of supporting information: author and cited work.
- Concept layer, which contains semantic description of the two essential categories – transformations (algorithms) and indicator classes (data entering or being exchanged between the algorithms)

[1] "A thing is... properly an indication if and where it in fact serves to indicate something to some thinking being... a common circumstance [is] the fact, that certain objects or states of affairs *of whose reality someone has actual knowledge indicate to him the reality of certain other objects or states of affairs, is the sense that his belief in the reality of the one is experienced ... as motivating a belief or surmise in the reality of the other*." [Italics by E. H.].

- Implementation layer, which contains information about the specific implementation of transformations and of validations of indicator classes in computer programs.
- Review/trust layer, containing user reviews and trust statements.

In our present implementation, each description on the conceptual layer consists of four basic elements:

- Human semantics: collection of human readable description, which helps the user to understand the meaning of the entity
- Metadata: pieces of computer recognizable data, which can be utilized e.g. to construct user interfaces
- Relations: relations of this entity to other entities
- Classifications: special case of relations that position the entity into external classification and terminology systems (UMLS, MeSH descriptors etc.).

Based on these principles, we have built a proof-of-concept implementation of a tool for semantics-based matching of data and algorithms. Based on the concepts described above, we dissected significant parts of the software "Compendium of Paediatric Auxology 2005" into a set of MediGrid indicator classes, transformations and their implementations. The software (MediGrid tool and the Paediatric Auxology example) is available at SourceForge; a pilot implementation on a web server is being used for testing and further development of the software.

In certain aspects, the Medigrid approach is very close to openEHR; we further analyzed similarities and differences.

3.2 MediGRID and OpenEHR

Both openEHR and MediGrid share some common features; they are based on the same principles, e.g. the "two-level" modeling of information, and both of them utilize phenomenological approach [3] in the description of the clinical data. The observable or computable clinical data are described as *Indicators* in MediGrid or *Observations* in openEHR; the grouping of the data in *Contexts* (MediGrid) is paralleled by the archetypes *Cluster* and/or *Folder* in openEHR. Due to this intrinsic compatibility, we are able to utilize the MediGrid methodology to process openEHR-encoded data. From the strict point of view, MediGrid entity models and entities fall into the archetype description, as defined in [4].

Carrier of the semantic information in the openEHR is the Details section, containing the definitions of use, misuse, copyright statement and the original resource URI. The mapping of the openEHR concepts to the entities required to fully semantically describe algorithms (utilized within the framework of MediGrid) is contained in Table 2. In order to utilize biomedical data bound to openEHR archetypes, the semantic constraints of both systems should match. The example of the concept of Body height, presented in openEHR and MediGrid (in two knowledge domains) is shown in Table 3. Similar tables could be constructed for other archetypes existing both in the openEHR and in MediGrid.

Table 2. Key elements of openEHR and MediGrid

Element	MediGrid	openEHR
User (reference to the author, reviewer, user of the software, implementer etc.)	User	Text encoded within each archetype
Existing (e.g. published) data reference available for peer review	Reference	Text encoded within each archetype
Observable (or computed) clinical data model (e.g. the body weight or body height)	Indicator class	Observation archetype
Hierarchical structuring mechanism (e.g. population ► patient ► exam ► E.N.T. exam)	Context	Cluster or Folder archetype
Data processing algorithm or procedure (for automatic or human application, e.g. BMI)	Transformation	Not implemented
Description of the algorithm's specific implementation	Implementation	Not implemented
Description of the accepted data	Validation	Encoded in the definition in each archetype
Mechanism for peer reviewing and trust statements	Review	Missing. Not required?

Table 3. Comparison of the Body height constraints

Declaration	openEHR	MediGrid: pneumology	MediGrid: anthropometry
Purpose	For recording the height or length of a person at any point in time, and in addition tracking growth and loss of height over time.	For application in the pneumological algorithms	For application in the anthropometrical algorithms, including the tracking of growth and growth dynamics
Constraints	Position – lying or standing Not to be used for growth velocity Difference from birth length Not estimated or adjusted	Position: standing Difference from birth length Not estimated or adjusted	Body height is defined as only in standing position (Difference from birth length) [Designed for use in calculations of growth velocity]
Measurement	The length of the body from crown of head to sole of foot.	Any allowed	Precise measuring method described

From the 166 currently existing archetypes described in openEHR and 94 entities of MediGrid implemented in the Faculty Hospital in Prague Motol, there are 10 concepts occuring in both systems. With the only exception (patient sex, which is required by e.g. the anthropometrical or pneumology algorithms and which is moved from the openEHR archetypes into the demographical section of the openEHR record), the data described in openEHR archetypes could be presented as MediGrid indicators and processed by the algorithms from the growing MediGrid library.

Another important difference is in the process of version checking, where the openEHR utilizes the versioning with backwards compatibility, whereas the MediGrid utilizes the quality control procedures and changes propagation.

4 Discussion

4.1 Indicator Ontology

The term biomedical ontology often refers to structuring the biomedical domain knowledge, e.g. disease taxonomies, medical procedures, anatomical terms, in a wide variety of medical terminologies, thesauri and classification systems. Besides various ontologies that describe the domain of anatomy, the most known of these systems are ICD, SNOMED-CT, GO, MeSH and UMLS, consisting usually of a thesaurus of biomedical terms or concepts on one side and a set of relationships between them on the other side. The basic hierarchical link between the thesaurus elements in most of these systems is the "is_a" relationship. If one element "is_a" another element then the first element is more specific in meaning than the second element. Other widely used relationships are "part_of", "result_of", "consist_of" or "associated_with". These relationships are perfectly suitable for describing basic semantic, functional or topologic structures. For more complex relations between concepts, as e.g. those represented in the domain of paediatric auxology, the ontology of indicators, indicator classes and transformations is much more productive: it allows to define and describe the key concepts and relations that constitute the domain, irrespective of their position in any general terminology / classification / ontology system. At the same time, if desired, relations to these "external" systems can be introduced in order to support mapping of concepts between domains.

4.2 Archetype Data as Indicators

Introducing the concept of indicators, indicator classes and indicator transformations, available for user assessment and subject to user decisions, into the archetype model of domain representation may contribute to interoperability of data and knowledge between different domains (represented by different archetypes). In fact, the use of explicitly documented indicator classes and transformations as archetype elements may help to maintain the archetypes as generic domain knowledge representation while allowing to process specific data (originating in different domains) and to use specific parts of that knowledge.

5 Conclusions

The need for proper support of traditional ways of medical information handling leads to specific requirements on the ways biomedical data and algorithms and their semantics are documented. Theoretical principles for implementation of a semantic-based tool for representation and application of biomedical knowledge have been postulated and verified by a non-trivial practical implementation. These principles may serve as a further step towards empowerment of users for better control over the knowledge that is represented in the ICT tools they are using.

Algorithms described in MediGrid notation can be directly used on data compliant with openEHR specifications (taking into account some specific issues). In such a way, data storage and transport, which is the main purpose of EHR systems, can be complemented with processing of the data in biomedical algorithms.

Acknowledgements

Supported by Czech research projects MediGrid, 1ET202090537 and by VZ FNM, MZO 00064203.

References

1. Fernandez-Breis, J.T., Menarguez-Tortosa, M., Martinez-Costa, C., Fernandez-Breis, E., Herrero-Sempere, J., Moner, D., Sanchez, J., Valencia-Garcia, R., Robles, M.: A Semantic Web-based System for Managing Clinical Archetypes. In: Conf. Proc. IEEE Eng. Med. Biol. Soc. 2008, pp. 1482–1485 (2008)
2. Vejvalka, J., Lesny, P., Holecek, T., Slaby, K., Jarolimkova, A., Bouzkova, H.: MediGrid - Facilitating Semantic-Based processing of Biomedical Data and Knowledge. In: Karopka, T., Correia, R.J. (eds.) Open Source in European Health Care:The Time is Ripe, pp. 18–21. INSTICC Press, Porto (2009)
3. Husserl, E.: Logical Investigations, Investigation I (translated by J. N. Finday), p. 184. Routledge, London (2001)
4. Beale, T.: Archetypes and the EHR. Stud. Health Technol. Inform. 96, 238–244 (2003)
5. The medical algorithms project, `http://www.medal.org`
6. Stroud, S.D., Erkel, E.A., Smith, C.A.: The use of personal digital assistants by nurse practitioner students and faculty. J. Am. Acad. Nurse Pract., 67–75 (2005)