Data and Knowledge for Medical Decision Support
B. Blobel et al. (Eds.)
IOS Press, 2013

doi:10.3233/978-1-61499-240-0-3

Knowledge Representation and Management Enabling Intelligent Interoperability – Principles and Standards

Bernd BLOBEL[1]
eHealth Competence Center, University Hospital Regensburg, Regensburg, Germany

Abstract. Based on the paradigm changes for health, health services and underlying technologies as well as the need for at best comprehensive and increasingly automated interoperability, the paper addresses the challenge of knowledge representation and management for medical decision support. After introducing related definitions, a system-theoretical, architecture-centric approach to decision support systems (DSSs) and appropriate ways for representing them using systems of ontologies is given. Finally, existing and emerging knowledge representation and management standards are presented. The paper focuses on the knowledge representation and management part of DSSs, excluding the reasoning part from consideration.

Keywords. Knowledge representation, decision support systems, artificial intelligence, system theory, architecture, ontologies, standards

Introduction

Increasing quality and safety in health and improving the efficiency of the care process requires best of breed solutions in a distributed, interoperable environment. Interoperability depends on motivation, willingness, interest, ability and skills to cooperate for meeting common business objectives [1]. The components of the solutions have to support the objectives of patient-centered health service delivery systems, thereby adapting their structure, function and interrelation according to the business needs, the necessary business processes, and contextual conditions. Here, the subject of care status and his/her preferences, environmental implications including natural, organizational, social ones, etc., must be taken into account. The management (definition, observation/measurement, interpretation) of the described factors requires knowledge, skills and experiences of the actors involved, which have to be shared either a priori or during the process in question. The level of a priori shared knowledge, skills and experiences defines the needed process-accompanying exchange and therefore the required interoperability level from an information perspective or the possible interoperability level from an organizational perspective. The technology evolution enabling new health paradigms such as pervasive health including mobile health require to include non-human actors in the consideration, leading to the term

[1] Corresponding Author. Bernd Blobel, PhD, Professor, Head; eHealth Competence Center, University of Regensburg Medical Center, Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany; Email: bernd.blobel@klinik.uni-regensburg.de; URL: www.ehealth-cc.de

principals (persons, organizations, devices, applications, components) defined by the Object Management Group in 1996 [2]. The aforementioned business process reengineering leads to distributed decision support systems, thereby changing the knowledge management models, which gets distributed as well [3]. Table 1 describes different interoperability levels from both perspectives.

Table 1. Interoperability levels from both an information and an organizational perspective

Information Perspective		**Organizational Perspective**
Interoperability Level	***Instances***	***Interoperability Level***
Technical interoperability	Technical plug&play, signal- & protocol compatibility	Light-weight interactions
Structural interoperability	Simple EDI, envelopes	Information sharing
Syntactic interoperability	Messages, clinical documents, agreed vocabulary	
Semantic interoperability	Advanced messaging, common information models and terminology	Coordination
Organizations/Service interoperability	Common business process	Collaboration Cooperation

Any approach to complex and intelligent health services considering multiple domains in the sense of translational medicine must be performed following a set of principles based on specific methodologies. The methodologies introduced in this paper are borrowed from system theory and systems engineering, modeling and good modeling practice, language theory, logics, and ontology engineering.

The paper first introduces definitions relevant to the topic addressed. Thereafter, we show how medical processes can be modeled using a system-theoretical, architecture-centric approach. The representation of clinical reality leads to the categorization of real world systems using established ordering systems in medicine. For the different phases of the development process as well as the different levels of expressivity, existing and emerging projects and standards for knowledge representation (KR) will be discussed. Decision Support Systems (DSSs) consist of three basic components: a knowledge base, an inference or reasoning engine, and means to communicate with the user. The paper focuses on the knowledge representation and management challenges and does not address the reasoning part and user interfaces of medical DSSs. Thereby, knowledge management is mainly considered as usual from a technological, but also from a cognitive perspective [4]. We also do not address non-knowledge-based DSSs. These are systems using machine learning as a form of Artificial Intelligence (AI) and/or pattern recognition in clinical data. Artificial neural networks and genetic algorithms are types of non-knowledge-based systems [5]. Finally, the paper does not cover knowledge base implementation aspects.

In the US, Clinical Decision Support Systems (CDSS) have been pushed as resonance of the Institute of Medicine report “To Err Is Human” [6] to improve patient safety.

1. Methods

For tackling the challenge addressed by the EFMI Special Topic Conference 2013 titled “Data and Knowledge for Medical Decision Support”, the key terms and concepts have to be defined as a common basis of communication and cooperation. DSSs are artifacts in the field of Artificial Intelligence (AI) tackling the formalization of ways of thinking, understanding, and problem solving taken by human beings. DSSs enable problem solving based on two things: knowledge and the ability to reason [7]. In general, AI can be considered from a) an input and operational perspective, b) an output and behavioral perspective, c) by comparing its fidelity with human performance, and d) by assessing its rationality, i.e., its ideal performance [8]. These different approaches result in different methodologies to model the system under consideration.

For managing complex environments such as the medical or even broader the health-related ones, the paper at hand follows a system-theoretical, architecture-centric approach. The feasibility of that approach has been demonstrated in several projects, specifications and standards already (e.g. [9]).

1.1. General definitions

In general, a *system* is a composition of interrelated elements, ordered in some way. Ordering schemas or categories in the system‘s context are: difference, identity, property or relation, and thing [10]. Another category system is: system, structure, element, and function [10]. We have used a combination of those definitions by stating: A system is a grouping of structurally and/or functionally interrelated components, which are separated from the components defining the environment by system boundaries. Systems can be composed (aggregated) to super-systems or decomposed (specialized) to sub-systems. They interact with their environment.

The *architecture* of a system describes its components, their functions and relations. Cooperating sub-systems form an interoperating system. We can therefore distinguish constructive or structural and behavioral or functional aspects of systems and their interrelated components. The quality of input and output could belong to one of the three categories: material, energy, or information [10]. Figure 1 shows input-output relations of systems [10].

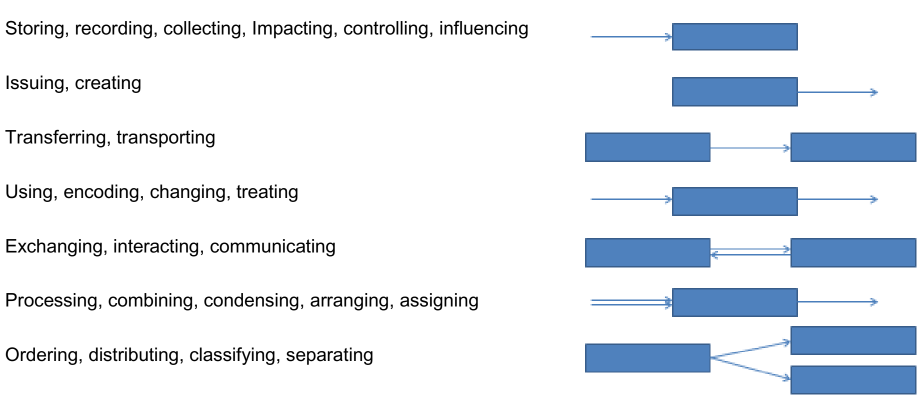

Figure 1. Input-output relations of systems [10]

There are several factors that make the description of living systems harder than describing technical artifacts. Here, metabolism with the attributes of self-organization and self-regulation as well as growth as consequence, reproduction (multiplying) with the attributes heredity (structure preservation) and mutation (structural change), and higher development through selection of best-adapted variants out of a large number have to be mentioned [10]. Another problem is the system's complexity, as described in Table 2 in equivalence with the translational medicine or systems medicine paradigm.

Table 2. Structural hierarchy in biology

Complexity	**Examples**
Chemical elements	H, C, N, P, O, S
Anorganic compounds	H_2O, CO_2, NO_3-, SO_4--, PO_4---
Organic basic elements	Amino acids, nucleic bases, fat acids, carbon hydrates
Macro molecules	Proteins, nucleic acids, poly saccharides, lipids
Cell organelles	Membranes, nuclei, mitochondria, ribosomes
Cells	Single-celled organisms, muscle cells, nerve cell (neuron)
Organs	Lung, stomach, brain, eye
Organisms, individuals	Kinds
Population	Groups of individuals, interrelations of individuals
Society	Social order

Taking cells as an example, there a many specializations such as metabolic cells, supporting cells, reproduction cells, moving cells (e.g. muscle cells), secretion cells, information processing cells, energy transformation cells. Among cell organelles, we have to distinguish endoplasmic reticulum, Golgi-apparatus, mitochondria, plasts, e.g. leuco-, chromo-, chloroplasts, lysosomes, e.g. microsomes, cytosomes, micro-tubuli and centrioles, ribosomes, vacuoles.

For tackling the complexity and evolution of living systems, the system's representation must be simplified by a *model* representing reality partially. A model is restricted to attributes the modeler is interested in. Defining the pragmatic aspect of a model, the interest is depending on the addressed audience, the reason and the purpose of modeling reality. Sometimes, the resulting model is used for a certain purpose and for a certain time instead of the original. Therefore, the model as a result of an interpretation must be interpreted itself [11]. From the representational perspective, a model is a statement expressed in a certain language [12]. Models are, therefore, propositions that may be either verbal, mathematical, or graphical (such as the Unified Modeling Language – UML) [13], in which entities are related according to the rules specified in a particular language. It provides an efficient and systematic way of representing knowledge about a system of interest and identifies missing information or necessary components not included yet. A model enables the prediction of system's behavior including the instantiation (values) of not accessible structural and/or functional facts by theoretical investigations (e.g. simulation), thereby testing hypotheses.

1.2. Definitions from the decision support systems' perspective

In the context of a decision support system, i.e. an interactive information system that provides information, models, and tools helping human professionals to take decisions, data are results of observations, measurements, and facts on which a decision as a choice or judgment made about something is to be based (adapted from [14]). Knowledge is a combination of instincts, ideas, rules, and procedures that enable the interpretation of data into information and that guide actions and decisions [15]. A KR system enables the logical interpretation of sentences in order to derive inferences from them. While the aforementioned definitions are ones the scientific community has at least some common views on, the definition of intelligence is quite shaky. The Latin root defines it as understanding, or the basis for choices, so getting closer to the definition of knowledge. In the addressed topic's framework, Alter defines intelligence as the first phase of the decision making process, covering the collection of information by scanning the environment and detecting problems to be solved [15]. Combining technical and cognitive aspects, intelligence can be explained by interpreting its basic principles data, information, knowledge, and wisdom [16]

1.3. Definitions from an information cycle's perspective

Selecting a particular business case as the system of interest, the information cycle can be used to interrelate concepts presented before. Considering the information cycle, *data* are facts, images or sounds that may be relevant in a business case. Data about the environment and the business case that intends to achieve the business objective is gathered to be entered into the system. This data is interpreted into information by a process of formatting, selecting and interrelating it, thus giving the data meaning. The information is used to derive decisions and to take actions. Decisions and actions and their consequences for the business case in its environment are assessed and evaluated with respect to the business objective. Knowledge is used to carry out formatting, selection, and interrelation of data, to appropriately derive decisions and to take actions, and to evaluate the outcome, based on the basic relations between the system's components according to Figure 1. The aforementioned statements are based on the different fundamental information definitions by Shannon, Brillouin, and Wiener related to the cycle phases data, information and action [17].

2. Results: Principles of Knowledge Representation

In the following, we will introduce principles and solutions for KR both in a formal and semi-formal way. In addition, we will give examples relevant to our scope of interest.

2.1. A system-theoretical, architecture-centric approach to knowledge representation

Design and deployment of DSSs meeting the knowledge-based interoperability challenge requires an architectural framework to model abstract systems and their instantiations. The Generic Component Model (GCM), which is applied successfully in a series of international projects, specifications, and standards, is such a framework [18-23]. The GCM is capable to describe the architecture of any systems, i.e. the composition and decomposition of its components. It allows multi-disciplinary considerations, i.e. the representation of different perspectives (partial systems or domains) of a system as established by domain experts using domain-specific terminologies and ontologies. In the context of intelligent system design it has been accepted that reasoning becomes simpler if the structure of the representation reflects the structure of the portion of reality being reasoned about [7, 24, 25]. Thereby, the representation of the GCM components, the structured objects, and their behavior, the processes KR deals with, must be mastered. In other words, the GCM is also used for modeling representation systems such as language, ontology, or in our special case, KR.

2.2. Data and Knowledge Representation in Medicine – Biomedical Ontologies

An *ontology* is a representation of the hierarchy of entities and the relations between the entities in a domain [26]. Reflecting this definition at the GCM results in the statement: An ontology provides the representation of a domain-specific architecture. Beside the philosophical definition of an ontology and especially in the context of DSSs, computer sciences defined "An (computational, *added by the author*) ontology is a formal, explicit specification of a shared conceptualization." [27]. Ontologies are expressed in logic-based formalisms, which provide (meta-) definitions of classes (concepts), relations, instances and axioms [28]. Reasons for developing ontologies are the need to sharing a common understanding about the structure of information between persons and software agents, the wish to enable re-use of domain knowledge, making domain assumptions explicit, separating domain knowledge and operational knowledge, and finally analyzing domain knowledge [29]. In the following, the evolution of knowledge representation in medicine will be considered in some more details, not claiming a comprehensive discussion.

After developing an abstract system's architecture, it must be instantiated for concrete business domains. The resulting real system components must be properly named and described, using pre-existent terminologies and ontologies where possible.

A *terminology* is the collection of all terms describing a certain domain. Observation units are terms, having a complex structure. A *term* describes properties of a real world object, thereby pointing out to that object and its properties. It may contain one or more modifiers. The construction of terms implies knowledge. Therefore, the classification of term sets is a trial to order complex knowledge. The underlying model is complex as well. It consists of characteristics and their molding (semantic concepts, terms), relations between terms of a terminology, and the classification for explicitly representing their relation. However, biomedical terminologies do not use formal and well-defined descriptions; they rather define the terms (if ever) by human language expressions, and express the associations between terms by informal, close-to human language relations [28].

A *classification*, or *systematics* is a systematic collection of abstract classes (concepts, types, or categories), which are used for dissociation and ordering. The single classes are commonly created by classification, i.e., by organizing objects according to certain characteristics and ordering them hierarchically. The set of class names forms a controlled vocabulary. The application of a classification to an object by selecting the appropriate class out of a given classification is called classification.

From the basic principles, two classification structures can be distinguished: mono-hierarchical and poly-hierarchical ones. In a *mono-hierarchy* (also called strong hierarchy, hierarchy with single inheritance), every class has only one parent class. The entire classification forms a tree structure. In a *poly-hierarchy* (also called weak hierarchy, hierarchy with multiple inheritance), a class can be derived from many parent classes. If a poly-hierarchy is highly distinctive and further relations between the classes are added, we call the outcome a *thesaurus*. The species assignment in biology is, e.g., called systematics.

Another classification structure separates *analytical classification* (from general to special, addressing pre-coordination) and *synthetic classification* (from special to general, addressing post-coordination). Most classifications are analytical ones. A prominent example for synthetic classifications is the *facette classification*.

A terminology is primarily a finite numeric list of terms, used to transfer information unambiguously. SNOMED (the Systematized Nomenclature of Medicine) is more than a general terminology as it connects terms and their codes with the ability to interrelate the terms in a meaningful way, maturing with SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) towards a medical ontology [30]. A clinical terminology encodes the complete description of clinical scenarios. So, a consultation can result in 38 codes. It is used by physicians for documenting patients' care (e.g. for establishing EHRs). The SNOMED CT core terminology provides a common language, which offers a consistent way for indexing, storage, retrieval, aggregation und communication of clinical data and facts between specialties and health organizations. Other examples for medical terminologies and ontologies are Logical Observation Identifiers Names and Codes (LOINC®) [31], OpenCyc [32], International Classification of Diseases (ICD) [33], the Medical Subject Headings (MeSH) [34], the Gene Ontology (GO) [35], the Generalized Architecture for Languages, Encyclopedias and Nomenclatures (openGALEN) [36], the Foundational Model of Anatomy (FMA) [37] and, for harmonizing between different concept representations, the Unified Medical Language System (UMLS) [38] and the OBO Foundry's Open Biological and Biomedical Ontologies [39, 40].

There are different application domains for medical classification and terminologies. For enabling structured documentation and communication in health care (telematics), a standardized transfer of observation results (LOINC), unambiguous labeling of diseases, health disorders and medicinal products/drugs (ICD-10-GM, Alpha-ID, SNOMED CT, ATC, ICF, …), as well as unambiguous identification of objects (OID), such as Doctor's letters/reports, hospitalizations, reha-memos, order-entry, or AMTS, ePrescription have been introduced. For indexing information, Medical Subject Headings (MeSH) have been defined to construct, e.g., literature databases or to retrieve information. The abbreviations used are explained at the end of the paper.

2.3. Basics of knowledge representation

The representation and communication of something is a matter of language. A *language* is a set of words composed of letters out of an alphabet. The language is defined by a grammar (formation rules) over that alphabet. Terms of natural languages have semantics, i.e. meaning and rules are expressed implicitly in the terms and their relations. Contrary, formal languages do not have semantics. They are often used as the basis for richer constructs endowed with semantics. Furthermore, they can be used to represent the syntax of formal theories. The system of logics belongs to the formal language family with further constructs, like proof calculi, which define a consequence relation.

Symbols, operators, and interpretation theory give sequences of symbols meaning within a KR. A key parameter in choosing or creating a KR is its *expressivity*. The more expressive a KR is, the easier and more compact it is to express a fact or element of knowledge within the semantics and grammar of that KR. However, more expressive languages are likely to require more complex logic and algorithms to construct equivalent inferences. A highly expressive KR is also less likely to be complete and consistent. Less expressive KRs may be both complete and consistent.

This property results in the complexity problem of formal language and reasoning systems with the lack of computability, at the same time losing the consistency of the language system. Natural languages are not only efficient in representing meaning, shared knowledge, skills, and experiences assumed. They provide an optimum between restriction to special structure and generative power enabling the rich and nevertheless sufficiently unambiguous representation of real world concepts, supported of course by common sense knowledge. This is one of the reasons for investing in natural language processing and not only relying on the formal representation of medical facts. Figure 2 provides an overview on KR languages or ontology types, evolving from informal KR representation languages up to formal ones. Regarding the aforementioned formalization and expressivity of more abstract, explicit ontology types, formal languages could be defined without restriction as it could be done with an unrestricted Turing Machine what is useless for the intended purpose of expressing natural language concepts, as it allows expressing everything without being complete and consistent. In the other extreme, a formal language generated with a highly restricted Markow Process is not useful, as it can only express quite simple concepts. Finally, we should not forget other languages such a graphical ones (Unified Modeling Language – UML) or even such exotics such as body language, all of them meeting the aforementioned principles.

To get closer to the vision of comprehensive interoperability, the ontological representations used by different domain experts for representing entities in reality must be harmonized. For that purpose, the ontological representation must be provided at a level of formalization and expressivity which guarantees common understanding, i.e. expresses meaning and rules as explicit as needed depending on education, skills, and experiences of the actors involved (Figure 2).

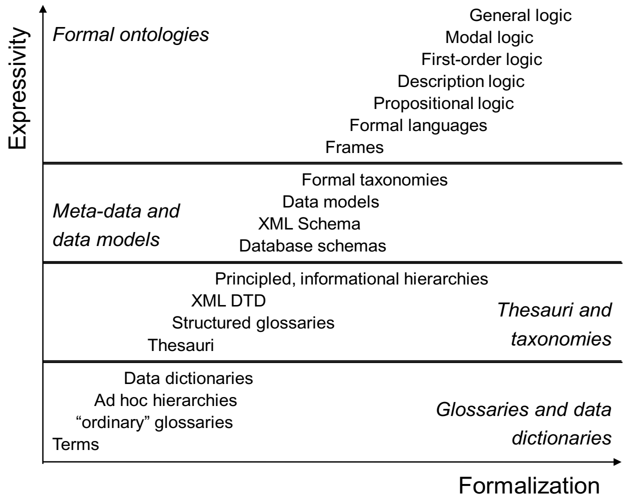

Figure 2. Types of ontologies

Originated in cognitive sciences, there are several KR techniques such as frames, rules, tagging, and semantic networks. Since knowledge is used to achieve intelligent behavior, the fundamental goal of KR is to facilitate reasoning, inferring, or drawing conclusions. A good KR has to manage both declarative and procedural knowledge.

Recent developments in KR include the W3C concept of the Semantic Web [41], and development of XML-based KR languages and standards, including Resource Description Framework (RDF) [42], RDF Schema for describing ontologies [43], Topic Maps [44], DARPA Agent Markup Language (DAML) [45], Ontology Inference Layer (OIL) [46], and Web Ontology Language (OWL) [47].

KR is first of all a surrogate for the thing itself to enable an entity to determine consequences by thinking (reasoning about the world) rather than acting. KR is a set of ontological commitments to answer the question about the terms to be used to think about the world. KR is a fragmentary theory of intelligent reasoning, expressed in terms of three components: the representation's fundamental conception of intelligent reasoning; the set of inferences the representation sanctions; and the set of inferences it recommends. KR is a medium for pragmatically efficient computation of thinking and a medium of human expression/language to describe the world (after Davis, Shrobe, and Szolovits [48]).

There are purpose related KR model types such as diagnostic models, connotative models, selective models, analytic models, instructive models, constructive models, or hybrid models.

Knowledge bases may represent inherent rules using set theory, Boolean logic, probability, Bayes rules, or informal logic according to the quality of relations of components and the strategy of the reasoning engine [49].

2.4. The system of ontologies

The system of ontologies (Figure 3) consists of one or more top level ontologies, from which the domain ontologies have to be derived, by that way enabling cross-domain harmonization. Examples for top-level ontologies are BFO [50], Cyc (and the non-commercial version OpenCyc), DOLCE [51], GFO [52], PROTON [53], and SUMO [54], but also combination of top-level ontologies such as COSMO [55], or MSO [56] (for the abbreviations see the Annex). After analyzing existing top-level ontologies [57], we referred in our work in deriving and harmonizing biomedical ontologies to the Basic Formal Ontology (BFO) [58] and sometimes also to the Descriptive Ontology for Linguistic and Cognitive Engineering DOLCE [51]. To facilitate interoperability between business processes (applications), domain ontologies must be specialized into application ontologies. For developing ICT solutions supporting the business case in question, the computation independent business processes (Business View) must be transformed into conceptual models reflecting the RM-ODP [59] Enterprise View, which are transformed into information model, computational models, and thereafter into implementable specifications for deployment. The RM-ODP representation of real world businesses deploys ICT-ontologies.

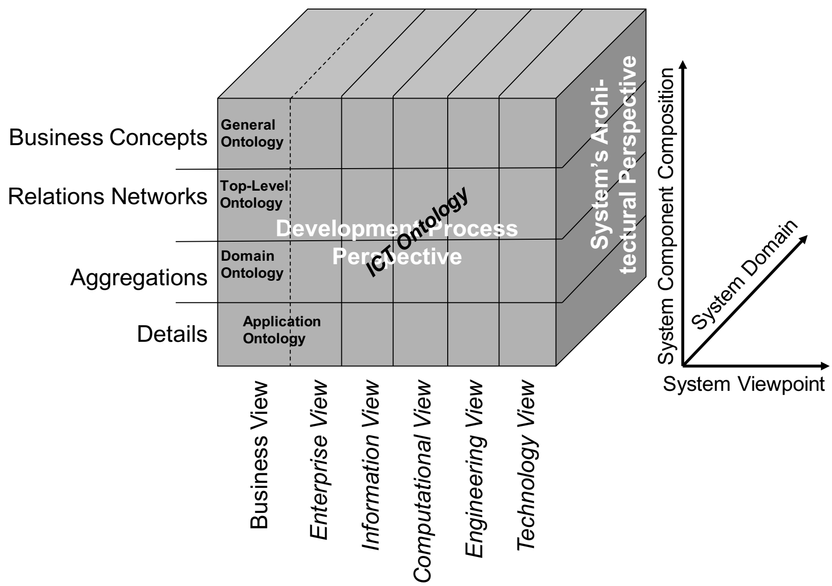

Figure 3. Modeling the system of ontologies using the GCM

A special ICT ontology is the well-known HL7 Reference Information Model (RIM) [60]. For the representation of the ICT solution space, the GCM can be deployed as well. So, the relation network of health domains tackled by the HL7 standards set is built by the Domain Model (or Domain Information Models – DIMs) derived from the RIM – not to be mixed up with the domain term referred to so far. At aggregations level, Refined Message Information Models (RMIMs) are developed as specializations of the DIMs, of course being RIM-based as well. At details level, the information components within a domain or domain-crossing Common Message Element Types (CMETs) are defined. Figure 3 demonstrates the use of the GCM to model the system of ontologies.

The system-theoretical, architecture-centric approach to DSSs based on the GCM framework enables formal system analysis, design, and management including the ontology harmonization needed for reasoning.

3. Results: Standards for Medical Knowledge Representation

For sharing computable clinical knowledge and enabling intelligent cooperation in distributed environments, a common language for specifying expressions and criteria is inevitable. Therefore, the aforementioned principles and solutions for KR must be standardized. This is a basic requirement for all presented levels of KR from the high level and generic up to domain- and application-specific ones, thereby also developing de-facto standards for corresponding tooling. Beside the basic standards tackling the challenge of KR, there are some health-specific ones addressed in the following. There are KR expression languages for guidelines representation and processing not considered in this paper because of the lack of international standardization. Here PRO*forma* [61, 62], Asbru [63, 64], EON [65, 66] have to be mentioned.

3.1. Arden Syntax

Arden Syntax has been developed for sharing medical knowledge stored in technically differently implemented knowledge bases. It could be called a technology-independent (or platform-independent) knowledge exchange format. The Arden Syntax represents this knowledge using frame logics. Arden Syntax encodes medical knowledge about individual decision rules in knowledge base form as self-contained Medical Logic Modules (MLMs), which can be embedded into proprietary clinical information systems. The MLMs are implemented as event-driven alerts or reminders.

Expressed as semiformal language, MLMs contain three slots or categories: the Maintenance Category (identifying the module, author, version, evidence level, etc.), the Knowledge Category (medical concept represented), and the Library Category (references/evidences). The knowledge category has a data slot on the one hand and evocation, logic, and action slots on the other hand. The latter specify the aforementioned events that trigger the evocation of the MLM, the logical criterion evaluated, and the action performed when the logical criterion is met. These knowledge-category components define the logical rule that the MLM specifies. The concept representation for describing medical conditions or recommendations contains a production rule and a procedural formalism, enabling a logical decision. Processes can be managed by chaining MLMs

Arden Syntax has been originally developed by New York Columbia Presbyterian Medical Center (CPMC) and IBM Health Industry Marketing in Atlanta, and thereafter wider used at Regenstrief Institute as well as within the HELP (Health Evaluation through Logical Processing) system at Salt Lake City LDS Hospital. Advanced applications using Arden Syntax for generating clinical alerts and reminders, interpretations, diagnoses, screening for clinical research studies, quality assurance functions, and administrative support in so-called event monitors are meanwhile globally deployed, as also demonstrated in this volume.

Arden Syntax has been standardized at first by ASTM (American Standards for Testing and Materials) [67] and thereafter at HL7 [60]. Since 2011, Arden Syntax

Version 2.8 is available, and a first version tackling fuzzy logic for production rule representation is under way. It is a specification compliant with the HL7 RIM.

3.2. GELLO

GELLO is a typed object-oriented standard query and expression language that provides a framework for management and processing of clinical data. Based on the OMG Object Constraint Language (OCL), GELLO enables the specification of decision criteria, algorithms and constraints on data and processes [68]. By that way, it provides a standardized framework for implementing DSSs. Therefore, GELLO is sometimes also called an object-oriented clinical decision support language [69].

The GELLO language can be used to build queries to extract and manipulate data from medical records and construct decision criteria by building expressions to correlate particular data properties and values. These properties and values can then be used in decision-support knowledge bases that are designed to provide alerts and reminders, guidelines, or other decision rules [60]. For this purpose, GELLO expresses logical conditions and computations in an standardized interchange format for modeling clinical guidelines, the GuideLine Interchange Format, v. 3 (GLIF3) [70]. Furthermore, it can be used for processing constraints, validation and calculated fields in Archetype data entries. It is also used to create complex data series for graphing or statistical analysis. For extracting data from any clinical database, a RIM-compliant virtual medical record has been defined as a mediator – similar to the RIM-based HL7 messaging framework enabling the communication of data between different health information systems. Thus, GELLO goes beyond the Arden Syntax which is limited to representing clinical rules.

GELLO is an HL7 International standard. Since 2010, GELLO Release 2 is available as formal, ANSI-approved specification. There are powerful GELLO compilers on the market, e.g., the Medical-Objects product [68].

3.3. GLIF

The GuideLine Interchange Format (GLIF) has been jointly developed at Stanford University, Brigham and Women's Hospital, and Columbia University to express and to share guidelines for prevention, diagnosis work-up, treatment, and patient-management processes (clinical pathways). They can be used as centrally stored sharable resource of knowledge, but also as directly executable guidance in response to network-based queries. Meanwhile, further institutions have joint the team.

GLIF3 [70] is an object-oriented expression and query language. Representing the description of complex multi-step guideline knowledge, the GLIF language can be also be translated into other languages established to execute clinical knowledge such as Arden Syntax. Using specific application interfaces (APIs), network-based clinical applications can directly access central decision support services executing approved guidelines based on the given data sets.

The GLIF3 specification consists of an extensible object-oriented model and a structured syntax based on the Resource Description Framework (RDF). GLIF3 enables encoding of a guideline at three levels: a conceptual flowchart, a computable specification that can be verified for logical consistency and completeness, and an implementable specification to be incorporated into local information systems. The GLIF3 model is represented using UML. Additional constraints are expressed in OCL.

For enabling the integration into information system, GLIF uses HL7 RIM classes and data types. While Arden Syntax follows a bottom-up approach vs the top-down approach of GLIF, both specifications are complementary for representing medical knowledge for clinical decision support.

GLIF3 is application independent, executable, can be easily integrated into clinical information systems, extensible, and offers a layered approach for managing the complexity of knowledge. It has been standardized at HL7 International. Corresponding tools have been developed, e.g., by the InterMed Collaboratory [71].

3.4. Archetypes

Based on specification provided by the EU project Good European Health Record (GEHR), Australia with Thomas Beale as main actor developed the Good Electronic Health Record (GEHR), which meanwhile evolves under the auspices of the openEHR Foundation [72].

The Archetype approach supports semantically enriched EHR systems by encapsulating the domain expert's knowledge in archetypes, defined and expressed using the Archetype Definition Language (ADL) [73]. ADL is a member of the OCL family. The Archetype model provides a constraint data model, thereby reflecting the domain experts' view. The structural Reference Model used is documentation specific, tackling storage and retrieval of information. Thus, it represents an informational perspective contrary to clinical facts described by translational medicine and sophisticated medical ontologies [74]. Using the Archetype Query Language (AQL) [75], clinical information can be consistently and easily retrieved with high improve recall and precision, thereby constraining the data object instances according to the Archetype definition. Archetypes represent clinical knowledge using frame logics. The Header part contains identifying information and meta-data including external ones. The Body part contains the clinical concept represented. The Terminology part finally references Archetype classes to standard terminologies, by that way supporting harmonization between different environments.

Archetypes and the Archetype approach have been standardized at ISO and CEN in the context of the ISO/CEN 13606 "EHR communication" standards series. openEHR offers freely available ADL parser [72].

3.5. HL7 Clinical Statement Model

HL7 International has developed the Clinical Statement Model for representing clinical concepts in a single message or document according to the HL7 Version 3 methodology. For sharing documented clinical information in a standardized way, HL7 developed the Clinical Document Architecture (CDA), representing clinical documents as structured, persistent, human-readable and machine-processable objects for a specific purpose. A CDA document consists of the CDA Header and the CDA Body. The latter contains information about CDA Structure, CDA Entries and CDA External References. HL7 v3 CDA documents and messages are encoded using the meta-language Extensible Markup Language (XML). They derive their machine processable semantics from the HL7 RIM and use the HL7 Version 3 data types and class structures, thereby providing a mechanism for the incorporation of concepts from standard coding systems such as SNOMED CT and LOINC.

In an evolutionary process, different levels of granularity for encoding information into machine-processable data have been defined, represented as different Releases of CDA. The CDA interoperability level enhances with more structured CDA Releases from R1 up to R3, as roughly explained in the following. In R1, just the Header has been fully specified, while the body is represented in just one block. In R2, the Body has been separated into tagged sections for diagnosis and treatment. In R3, the Body part will be structured up to the level of atomic concepts. HL7 Templates are a constraint on the CDA R2 object model and/or against a Clinical Statement Model [60].

3.6. The Clinical Information Modeling Initiative

The Clinical Information Modeling Initiative (CIMI) is an international action to provide a common format and a common development process for detailed specifications for the representation of health information to enable creation and sharing of semantically interoperable information in health records, messages, documents, and for secondary data uses. CIMI is mainly based on the aforementioned Archetype approach. Additionally to the Archetype Object Model [76] and the expression means of ADL[73], an extended Reference Model [77] and the representation of the entire development process using UML and the SOA framework will be deployed. For more information, the reader is referred to [74].

3.7. Security and privacy concept representation

Security and privacy concerns are essential in the context of communication of, and collaboration based on, personal health information. Security and privacy related knowledge specifying the underlying concepts as well as applicable rules and regulations are expressed in policies.

A system-oriented, architecture-centric definition of the policy system and its representation is given in [78, 79], and meanwhile standardized at ISO (e.g., ISO 22600 Health informatics – Privilege management and access control [9]). Partially reusing the PONDER policy language concept [80], the system of policies has been developed according to the GCM framework as well as based on meta-models. The concept representation of the policy system has been provided using UML, but also different logic languages.

Another way for formally modeling policies and ruling access control is the Extended Access Control Markup Language (XACML) [81], developed at OASIS with the XML meta-language. OASIS' Security Assertion Markup Language (SAML) [82] defines security services assigned to entities in a header-body-reference structure using XML. There are also less formal concept representations related to healthcare and security as, e.g., the HL7 RBAC approach, which is essentially a vocabulary effort [60].

4. Discussion

Medical DDSs require the presentation of medical facts, rules, and processes, i.e., the interpretation of observations and the derivation of reasonable actions, in a maschine-processable way. The paper has introduced in some detail the special challenge of developing, formalizing, representing, implementing, and processing the multi-disciplinary knowledge defining the healthcare and the health system. The paper has limited its focus to the first steps of the creation, formalization, representation, and implementation of knowledge in medicine.

After discussing the drivers for the development of automated decision support to improve safety and efficiency of health processes, the foundation for knowledge creation, management and representation in medically defined IT environment have been introduced. Here, especially the system-oriented, architecture-centric approach to solutions for medical decision support based on the Generic Component Model must be mentioned. Furthermore, definitions for the fundamental terms and concepts, such as interoperability, systems, architecture, model, language (at different level of formalization), knowledge, data, information, term, terminology, classification, systematics, ontology, and even the shaky concept of intelligence have been – sometimes controversially – introduced.

The framework clinical concept representation is based on can be informal, semiformal, or formal. We should have in mind that knowledge representation by ontology types with increased expressivity and formalization – i.e. making implicit concepts and relations explicit – implies the increase of complexity of that representation. Of course, some complexity might be hidden in the tooling deployed. Some important and frequently internationally standardized frameworks for representing clinical concepts in an informal, semiformal, or formal way, i.e., being based on vocabularies, terminologies, or ontologies are mentioned.

Finally, common and internationally standardized representations for sharing computable clinical knowledge and enabling intelligent cooperation in distributed environments are presented and shortly discussed. Here, Arden Syntax, GELLO, openEHR Archetypes, HL7 Clinical Statements, OASIS security and privacy concept representations, but also the Clinical Information Modeling Initiative artifacts, reusing and a little extending the archetype approach, are presented in some more or less details.

Table 3. Concept representation approaches [11]

Approach / **Components**	**GEHR/ openEHR**	**Arden**	**HL7 EHR/ CDA**	**SOA/MDA**	**OASIS security services**
Business modelling	Archetypes	Common language	Clinical Templates	TOGAF, CIM, MOF	--
Knowledge representation	GEHR parts	Arden Syntax Categories	HL7 CDA Structure	OCL Package	SAML/ XACML
Identification	Header (contains also ext. metadata)	Maintenance category	Header	Header	Header
Content	Body	Knowledge category	Body	Body structure, Body	Body
References	Terminology	Library category	Embedded terminology. Ext. refs	External reference	Ext. refs
Substructures	Blocks	Slots	Entries	UML components	Elements
Language	ADL	Semiformal language, logical ops.	XML	OCL, typed specification language	XML

In Table 3, the different approaches to clinical concept representation are systematically compared. GELLO as an OCL language comparable with ADL, OCL, or XML has not been included in that table. Instead, and despite of not tackling clinical concepts, the representation of concepts within OMG's Model-Driven Architecture (MDA) and its advancement to the Service-Oriented Architecture (SOA) is included in the table. The comparison demonstrates certain similarities, as all approaches use meta-languages as typed specification languages. They use types of frame logic to represent semantics and refer to accepted and standardized terminologies and ontologies.

A further discussion can be found in [11]. Readers more interested in reasoning are referred to [83]

Acknowledgement

The author is indebted to thank Mathias Brochhausen, University of Arkansas of Medical Sciences, Little Rock, Arkansas, U.S.A., for careful review and highly valuable recommendations. Furthermore, he thanks Jana Zvárová and Arie Hasman for their corrections and support.

References

[1] Blobel B. Architectural Approach to eHealth for Enabling Paradigm Changes in Health. Methods Inf Med 2010; 49(2): 123–134.
[2] Object Management Group, The CORBA Security Specification. OMG Doc.No. 95-12-01. Inc. http://www.omg.org. Last access: December 1, 2009.
[3] Kühn Pedersen M, and Holm Larsen M. Distributed knowledge management based on product state models – the case of decision support in health care administration. Decision Support Systems 2001; 31:139-158.
[4] Kurtyka J. The Cognitive Dimension of Knowledge Management. EIMI Archives 2008; 1(11).
[5] Berner ES, and La Lande TJ. Overview of Clinical Decision Support Systems. In: Berner ES (ed) Clinical Decision Support Systems – Theory and Practice. 2nd Edition. New York: Springer Science+Business Media; 2007.
[6] Kohn LT, Corrigan JM, and Donaldson MS (eds) To Err Is Human – Building a Safer Health System. Institute of Medicine. Washington, D.C.: National Academic Press; 1999.
[7] Szolovits P. Artificial Intelligence and Medicine. Chapter 1 in Szolovits, P. (Ed.) Artificial Intelligence in Medicine. Boulder, Colorado: Westview Press; 1982.
[8] Russell S, Norvig P. Artificial Intelligence – A Modern Approach. Boston; Pearson Education Inc.: 2010.
[9] International Organization for Standardization. ISO 22600 Health informatics – Privilege management and access control. Geneva; 2006. www.iso.org (last accessed on 20 February 2013).
[10] Völz H. Information. Berlin: Akademie-Verlag; 1982.
[11] Blobel B. (2006) Concept Representation in Health Informatics for Enabling Intelligent Architectures. Stud Health Technol Inform 2006; 124:285-291.
[12] Massoud TF, Hademenos GJ, Young WL, Gao E, Pile-Spellman J, and Vinuela F. Principles and philosophy of modeling in biomedical research. FASEB J 1998 Mar; 12(3):275-85.
[13] Object Management Group, Inc. Unified Modeling Language (UML). www.omg.org/spec/UML/ (last accessed on 20 February 2013).
[14] Collins Compact English Dictionary, New Editoin. Glasgow: Harper Collins Publishers; 1994.
[15] Alter S. Information Systems – A Management Perspective. Reading: Addison-Wesley Publishing Company; 1992.
[16] Pejman P. Methodic Applied Knowledge to Hyper Fictitious Intelligence, 2008 (MAKHFI.com). http://www.makhfi.com/ (last accessed on 20 February 2013).
[17] van Bemmel JH, Musen MA (eds) Handbook of Medical Informatics. Heidelberg: Springer; 1997.

[18] De la Cruz E, Lopez DM, Blobel B (2012) A Reference Architecture for Sharing Documents in Colombia. European Journal for Biomedical Informatics 8, 3; 2012: en11-en17.
[19] Vida M, Stoicu-Tivadar L, Blobel B, Bernad E (2012) Modeling the Framework for Obstetrics-Gynecology Department Information System. European Journal for Biomedical Informatics 2012; 8(3):en57-en64.
[20] Yildirim Yayilgan S, Blobel B, Petersen F, Hovstø A, Pharow P, Waaler D, Hijazi Y (2012) An Architectural Approach to Building Ambient Intelligent Travel Companions. International Journal of E-Health and Medical Communications 2012; 3(3):86-95.
[21] Ruotsalainen P, Blobel B, Seppälä A, Sorvari H, Nykänen P (2012) A Conceptual Framework and Principles for Trusted Pervasive Health. J Med Internet Res 2012; 14(2):e52. URL: http://www.jmir.org/2012/2/e52/ (last accessed on 20 February 2013).
[22] Lopez DM, Blobel B (2009) A development framework for semantically interoperable health information systems. International Journal of Medical Informatics 2009; 78(2):83-103.
[23] Bernal JG, Lopez DM and Blobel B (2012) Architectural Approach for Semantic EHR Systems Development Based on Detailed Clinical Models. Stud Health Technol Inform 2012; 177:164-169.
[24] Bobrow DG, and Winograd T. An Overview of KRL, a Knowledge Representation Language. Technical Report AIM-293, Stanford Artificial Intelligence Lab., Stanford, Ca; 1976. ftp://reports.stanford.edu/pub/cstr/reports/cs/tr/76/581/CS-TR-76-581.pdf (last accessed on 20 February 2013).
[25] Szolovits P, Hawkinson L, and Martin WA. An Overview of OWL, a Language for Knowledge Representation. In: Rahmstorf G, and Ferguson M (eds) Proceedings of the Workshop on Natural Language Interaction with Databases, International Institute for Applied Systems Analysis, Schloss Laxenburg, Austria, 10 Jan 1977.
[26] Guarino N. Formal ontology in information systems. In: Guarino N (ed) Formal Ontology in Information Systems, Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy, 3-15. Amsterdam: IOS Press; 1998.
[27] Studer R, Benjamins R, and Fensel D. Knowledge engineering: Principles and methods. Data & Knowledge Engineering 1998; 25(1–2):161–198.
[28] Freitas F, Schulz S, Moraes E: Survey of current terminologies and ontologies in biology and medicine. RECIIS - Electronic Journal in Communication, Information and Innovation in Health 2009; 3(1):7-18.
[29] Noy NF, and McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
[30] IHTSDO: SNOMED CT®. http://www.ihtsdo.org/snomed-ct/ (last accessed on 20 February 2013).
[31] Logical Observation Identifiers Names and Codes (LOINC®). loinc.org (last accessed on 20 February 2013).
[32] Cycorp. OpenCyc. http://www.cyc.com/platform/opencyc (last accessed on 20 February 2013).
[33] World Health Organization. International Classification of Diseases. www.who.int (last accessed on 20 February 2013).
[34] U.S. National Library of Medicine. Medical Subject Headings (MeSH) http://www.nlm.nih.gov/mesh/ (last accessed on 20 February 2013).
[35] The Gene Ontology (GO) Consortium. The Gene Ontology. http://www.geneontology.org/ (last accessed on 20 February 2013).
[36] openGALEN Foundation. The Generalized Architecture for Languages, Encyclopedias and Nomenclatures (openGALEN) www.opengalen.org/ (last accessed on 20 February 2013).
[37] University of Washington, Structural Informatics Group. Foundational Model of Anatomy (FMA). fma.biostr.washington.edu/ (last accessed on 20 February 2013).
[38] U.S. National Library of Medicine. Unified Medical Language System® (UMLS®). http://www.nlm.nih.gov/research/umls/ (last accessed on 20 February 2013).
[39] The OBO Foundry. The Open Biological and Biomedical Ontologies. www.obofoundry.org/ (last accessed on 20 February 2013).
[40] Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 2007; 25(11): 1251-1255.
[41] World Wide Web Consortium. Semantic Web. www.w3.org/standards/semanticweb/ (last accessed on 20 February 2013).
[42] World Wide Web Consortium. Resource Description Framework. www.w3.org/RDF/ (last accessed on 20 February 2013).
[43] World Wide Web Consortium. RDF Schema. www.w3.org/TR/rdf-schema/ (last accessed on 20 February 2013).
[44] TopicMaps.org Consortium. Topic Maps. www.topicmaps.org/ (last accessed on 20 February 2013).

[45] The DARPA Agent Markup Language (DAML) Program. DARPA Agent Markup Language (DAML). www.daml.org/ (last accessed on 20 February 2013).
[46] The DARPA Agent Markup Language (DAML) Program. Ontology Inference Layer (OIL). http://www.daml.org/language/
[47] World Wide Web Consortium. Web Ontology Language (OWL). www.w3.org/standards/techs/owl (last accessed on 20 February 2013).
[48] Davis R, Shrobe H, and Szolovits P. What is a Knowledge Representation? AI Magazine 1993; 14(1):17-33.
[49] Spooner SA. Mathematical Foundations of Decision Support Systems. In: In: Berner ES (ed) Clinical Decision Support Systems – Theory and Practice. 2nd Edition. New York: Springer Science+Business Media; 2007.
[50] IFOMIS. Basic Formal Ontology (BFO). www.ifomis.org/bfo (last accessed on 20 February 2013).
[51] EU Project WonderWeb: Ontology Infrastructure for the Semantic Web. A Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). http://wonderweb.man.ac.uk/ (last accessed on 20 February 2013).
[52] Onto-Med. General Formal Ontology (GFO). www.onto-med.de/ontologies/gfo/
[53] EU Project Semantically-Enabled Knowledge Technology (SEKT). PROTo ONtology (PROTON). proton.semanticweb.org/ (last accessed on 20 February 2013).
[54] IEEE. The Suggested Upper Merged Ontology (SUMO). http://www.ontologyportal.org/ (last accessed on 20 February 2013).
[55] COSMO Foundation. The Common Semantic Model (COSMO).
[56] IEEE. The Multi-Source Ontology (MSO). http://suo.ieee.org/ (last accessed on 20 February 2013).
[57] Uribe GA, Lopez DM and Blobel B. Architectural Analysis of Clinical Ontologies for pHealth Interoperability. Stud Health Technol Inform 2012; 177:176-182.
[58] BFO2: http://code.google.com/p/bfo/ (last accessed on 20 February 2013).
[59] International Organization for Standardization. ISO/IEC 10764 Reference Model Open Distributed Processing. Geneva; 1996 www.iso.org (last accessed on 20 February 2013).
[60] Health Level 7 International, Inc. www.hl7.org
[61] Fox J, Rahmanzadeh A. Disseminating medical knowledge: the PROforma approach. Artificial Intelligence in Medicine 1998; 14:157-181.
[62] PRO*forma*. www.openclinical.org/gmm_proforma.html (last accessed on 20 February 2013).
[63] Shahar Y, Miksch S, Johnson P. The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. Artificial Intelligence in Medicine 1998; 14:29-51.
[64] Asbru. www.openclinical.org/gmm_asbru.html (last accessed on 20 February 2013).
[65] Tu SW, Musen MA. A Flexible Approach to Guideline Modeling. Proc AMIA Symp 1999:420-424.
[66] EON. www.openclinical.org/gmm_eon.html (last accessed on 20 February 2013).
[67] American Standards for Testing and Materials (ASTM). www.astm.org (last accessed on 20 February 2013).
[68] Sordo M, Boxwala AA, Ogunyemi O, Greenes RA. Description and status update on GELLO: a proposed standardized object-oriented expression language for clinical decision support. Stud Health Technol Inform 2004; 107(Pt 1):164–168.
[69] Medical-Objects. GELLO.org. www.gello.org (last accessed on 20 February 2013).
[70] Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, Patel VL, Greenes RA, Shortliffe EH. GLIF3: A Representation Format for Sharable Computer-Interpretable Clinical Practice Guidelines. J Biomed Inform 2004 Jun; 37(3):147-61.
[71] The InterMed Collaboratory. http://mis.hevra.haifa.ac.il/~morpeleg/Intermed (last accessed on 20 February 2013).
[72] openEHR Foundation. www.openehr.org (last accessed on 20 February 2013).
[73] Beale T, Heard S. Archetype Definition Language ADL 1.5. openEHR Foundation, January 2012, www.openehr.org (last accessed on 20 February 2013).
[74] Blobel B, Goossen W, Brochhausen M. Clinical Modeling – a Critical Analysis. (submitted to International Journal of Medical Informatics)
[75] Beale T. openEHR Archetype Query Language Description. http://www.openehr.org/wiki/display/spec/Archetype+Query+Language+Description (last accessed on 20 February 2013).
[76] Beale T. Archetype Object Model AOM 2.1. openEHR Foundation, January 2012, www.openehr.org (last accessed on 20 February 2013).
[77] CIMI Reference Model Report, Draft V 0.3, May 2012
[78] Blobel B. Ontology driven health information systems architectures enable pHealth for empowered patients. International Journal of Medical Informatics 2011; 80:e17-e25.

[79] Blobel B, Nordberg R, Davis JM, Pharow P. Modelling privilege management and access control. International Journal of Medical Informatics 2006; 75(8):597-623.
[80] Damianou N, Dulay N, Lupu E, Sloman M. Ponder: a language for specifying security and management policies for distributed systems: the language specification, version 2.3. Imperial College Research Report DoC 2000-2001, 20 October 2000.
[81] Organization for the Advancement of Structured Information Standards (OASIS), XACML 2.0. http://www.oasis-open.org/specs/index.php#xacml (last accessed on 20 February 2013).
[82] Organization for the Advancement of Structured Information Standards (OASIS), SAML. https://www.oasis-open.org/standards#samlv2.0 (last accessed on 20 February 2013).
[83] van Harmelen F, Lifschitz V, Porter B (eds) Handbook of Knowledge Representation. Amsterdam: Elsevier; 2008

Annex

Table A1. Abbreviations not directly explained in the text

Alpha-ID	Number which identifies the entries in the alphabetical index of the ICD-10-GM
AMTS	Abbreviated mental test score
ATC	Anatomical-therapeutic-chemical classification
COSMO	Common Semantic Model, an upper ontology
GFO	General Formal Ontology
ICD-10-GM	ICD-10-German Modification
ICF	International Classification of Functioning, Disability and Health
MSO	Multi-Source Ontology
OID	Object Identifier
PROTON	Proto Ontology, an upper ontology
SUMO	Suggested Upper Merged Ontology