Semantic Web technologies for managing EHR-related clinical knowledge

Catalina Martínez-Costa[1], Marcos Menárguez-Tortosa[1], José Alberto Maldonado[2], Jesualdo Tomás Fernández-Breis[1]
[1]Universidad de Murcia, [2]Universidad Politécnica de Valencia
Spain

1. Introduction

The Semantic Web (Berners-Lee et al, 2001) is a vision of the future Web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web. Semantic Web technologies promise to be capable of facilitating the management of knowledge and promote semantic interoperability between systems. There are different basic technologies for the success of the Semantic Web, amongst which the cornerstone technology is the ontology. In literature, multiple definitions for ontology can be found (see for instance (Gruber, 1993; van Heijst et al, 1997)). An ontology represents a common, shareable and reusable view of a particular application domain, and they give meaning to information structures that are exchanged by information systems (Brewster et al, 2004).

Semantic Web technologies and, in particular, ontologies have been identified in the final report of the Semantic Health project (SemanticHealth Report, 2009) as one of the basic technologies for the consecution of semantic interoperability of healthcare information systems. In healthcare, interoperability refers to the ability of different systems and organizations to exchange information and to use the information that has been exchanged.

On the other hand, the medical field is continually changing, and new findings about diseases and clinical treatments are continually made. Huge amounts of heterogeneous medical information and clinical terms are generated. However, the standardization of clinical information and knowledge has not been researched until the 90s. Recently, different architectures for exchanging clinical information and knowledge have been proposed, and the dual model-based one seems to be the most promising. This standard architecture introduces a separation between knowledge and information where knowledge reflects the possible changes. This separation is carried out by means of a double model (Beale, 2001), the reference model and the archetype model. The reference model reflects the generic and stable properties of the electronic healthcare record, whereas the archetype model represents the knowledge level, and consists of clinical concepts, called archetypes, that are based into entities of the corresponding reference model.

Hence, the methodology for the development of health information systems is changing and the dual model approach proposes a semantic layer defined by the archetypes. The

semantics in archetypes have a double nature: structural and terminological. By structural, we mean that the proper structure of the archetype provides some semantics. In addition to this, an archetype can be seen as a set of interrelated conceptual clinical entities. Each entity has a set of terminological bindings associated, which are specified by means of links to terms of specific medical terminologies, such as SNOMED-CT (SNOMED-CT).
Many medical terminologies have been recently or are in the process of being represented in the Web Ontology Language (OWL), because its formal nature allows for a better management of clinical knowledge. Furthermore, the common representation of archetypes and terminologies in OWL would allow a uniform management of clinical knowledge, which would also facilitate the consecution of semantic interoperability.
Hence, in this chapter we describe how Semantic Web technologies can be used to manage such clinical knowledge, and has two main streams:

- Representation of clinical archetypes: Clinical archetypes can be represented as OWL ontologies. We will describe an approach that combines Semantic Web technologies and Model-driven Engineering (Douglas et al., 2006) to achieve the goal. This approach can be applied to any dual-model based EHR standard.
- Management of clinical archetypes: The management of clinical archetypes will be illustrated by describing an EHR-independent Semantic Web system for managing archetypes. This system allows for annotating archetypes with external resources, performing searches and classification tasks.

2. Electronic Healthcare Records

Health information systems from hospitals and primary care organizations are expected to be capable of communicating to support the continuous medical process of the patient at local, regional, national and international level.
The Electronic Healthcare Record (EHR), defined as a repository of information regarding the health of a subject of care, in computer processable form (ISO/TC 215 TR, 2008), constitutes the cornerstone technology for the achievement of that goal. Its primary purpose is to provide a documented record of care that supports present and future care by the same or other clinicians. Among other benefits, the replacement of the traditional paper-based patient records with EHRs will increase the quality and efficiency of the patient medical care and will cut back on costs.

Nowadays there are different advanced approaches as standards or specifications for representing and communicating EHRs such as HL7 (HL7), OpenEHR (OpenEHR), and EN 13606 (UNE-EN 13606).
HL7 stands for Health Level Seven, and was founded in 1987 to provide healthcare standards for the exchange, management and integration of clinical information. There are several HL7 implementations. It is worth pointing out HL7 v2.X, that is focused on the exchange of messages and which have been widely used in America and Europe by the industry. More recently HL7 v3 was proposed, introducing the Reference Information Model (RIM) and the Clinical Document Architecture (CDA). This last standard version has (ISO).
On the other hand, the European Health Record (GEHR) project (1991-95) contributed to develop the OpenEHR specification. Following GEHR several projects extended and refined

its results. All these ones influenced the creation of the OpenEHR specification by the non-profit organization OpenEHR Foundation (OpenEHR).
Finally, the EN 13606 standard was influenced by OpenEHR. EN 13606 has been drawn on the practical experience of its predecessor ENV 13606. In fact, it is considered a subset of the full OpenEHR specification oriented to the exchange of EHR extracts.
Both the OpenEHR specification and the EN 13606 standard share the same modelling architecture. This architecture is named dual model-based architecture and has influenced the HL7 v3 standard. It is explained in next section in detail since our work will be focused on the semantic management of dual model-based standards.

2.1 Dual model architecture

The main feature offered by this modelling architecture is the separation between information and knowledge. On the one hand, information is modelled by means of a reference model (RM) and on the other hand, knowledge is modelled using an archetype model (AOM). The first one is specific to the healthcare domain but still very generic. It defines the set of classes that forms the generic building blocks of the EHR and it is stable over time. *Person or clinical session* would be classes of this reference model. The second one represents healthcare and application specific concepts such as the *measurement of cholesterol*, *the blood pressure* and so on by using archetypes.
An archetype describes configuration of data instances whose classes are described in the reference model. They are defined using the Archetype Definition Language (ADL). This language provides a concrete syntax for expressing them as text documents. Figure 1 illustrates the relationships of archetypes with data. Archetypes are instances of an archetype model which is a common formalism for expressing all archetypes. The archetype object model (right side) is formally related to the reference model, such that its semantics are those of constraint on instances of classes defined in the reference model (left side). If data are created and modified using archetypes, archetypes constrain the configuration of data instances to be valid according to the archetype.

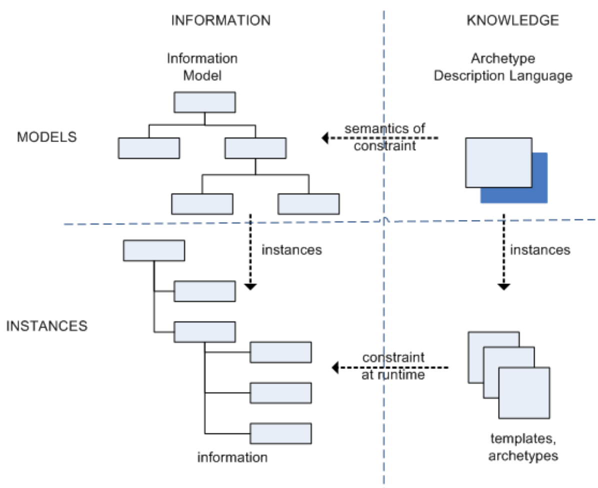

Fig. 1. Archetype Model Meta-architecture (Meta-Architecture)

Both OpenEHR and EN 13606 are based on the dual model architecture. However, they differ in how they structure the EHR domain, that is, they define different reference models. Thus, archetypes will be defined as constraints on these reference model classes for each standard and will be written in ADL.

Figure 2 shows an extract of an ADL archetype for the visual acuity recording for the EN 13606 standard. In the figure, the different ADL sections of the archetype can be observed: header, description, definition and ontology. The header includes the name of the archetype, specialization information and so on. In the Figure the header includes the name of the archetype *(CEN-EN13606-ENTRY.visual_acuity.v1),* the language it is written in *(ISO_639-1::en)* and the archetype concept code *(at0000)*. The description section includes audit information, such as original author, purpose or lifecycle status. The definition section contains the structure and restrictions associated to the clinical concept defined by the archetype. Here, this section says that visual acuity is recorded by means of a table, whose row head is “Left”, “Right”, “Both eyes” and whose columns are the following values ("5/6" ,"6/6", "6/7.5", "6/9", "6/12", "6/18", "6/36", "6/60", "Count fingers", "Perceive light", "Blind"). That is, it expresses the acuity value of each eye separately and of both of them together. Finally, the ontology section includes the terminological definition and bindings. In this last section the linguistic expressions associated to the terms from the definition part are provided, as well as their possible bindings in other terminologies. For instance, the link to the SNOMED-CT term says how visual acuity is defined in the SNOMED-CT terminology.

Archetypes combine to form templates. They usually correspond to screen forms, printed reports, and in general, complete application-level information to be captured or sent. They are generally developed and used locally, recording the specific needs of the user or institution, while archetypes are usually widely used.

In fact, archetypes may constitute a clinical guide for clinicians and its importance can be noticed in some acts as the adoption of the European EHR EN 13606 standard by Sweden for their national EHR developments (Swedish-decision).

```
archetype (adl_version=1.4)  CEN-EN13606-ENTRY.visual_acuity.v1
concept      [at0000]
language     original_language = <[ISO_639-1::en]>
description  ...
definition
  ENTRY[at0000] occurrences matches {1..1} matches {  -- Visual acuity
    items existence matches {0..1} cardinality matches {1..1; unordered} matches {
          ...
             CLUSTER[at0003] occurrences matches {1..1} matches {  -- Table
               parts existence matches {0..1} cardinality matches {0..1; unordered} matches {
                 CLUSTER[at0004] occurrences matches {0..1} matches {  -- row
                   parts existence matches {0..1} cardinality matches {2..2; ordered} matches {
                     ELEMENT[at0005] occurrences matches {0..1} matches {  -- row head
                       value matches {
                         SIMPLE_TEXT occurrences matches {1..1} matches {
                           originalText matches {"Left","Right","Both eyes"}
                         }}}
                     ELEMENT[at0006] occurrences matches {1..1} matches {  -- Visual acuity
                       value matches {
                         ORD occurrences matches {1..1} matches {
                           symbol matches {
                             CODED_TEXT occurrences matches {1..1} matches {
                               codedValue matches {
```

```
                                CD occurrences matches {1..1} matches {
                                   displayName matches {
                                      "5/6","6/6","6/7.5","6/9","6/12","6/18","6/36",
                                      "6/60","Count fingers","Perceive light","Blind"
                                }}}}}
                          value matches {5,6,7,9,12,18,36,60,100,200,300}
                       }}}}}}}...
ontology
   terminologies_available = <"SNOMED-CT", ...>
   term_definitions = <
      ["en"] = <
         items = <
            ["at0000"] = <
               text = <"Visual acuity">
               description = <"The functional acuity of vision, aided and unaided">
            >...
   term_binding = <
      ["SNOMED-CT"] = <
         items = <
            ["at0000"] = <[SNOMED-CT::363983007]>
         >>>...
```

Fig. 2. Extract of the visual acuity EN 13606 ADL archetype

2.2 ADL Limitations

Despite ADL is the language adopted for the archetypes description, it has some limitations. The parsing of an ADL archetype will return objects according to the Archetype Object Model (AOM). This model is common to all dual model-based standards, that is, it will have no information about the particular reference model for which the archetype has been built. Thus, the obtained objects cannot be used to perform any semantic activity, such as comparison, selection, or classification. Also, it does not provide any component that guarantees the consistency of the clinical knowledge but only at archetype level, that is, the conformance to ADL/AOM principles. Therefore, to process ADL content, two elements are needed: an ADL parser to get the AOM objects and the validator for the particular reference model to guarantee the clinical correctness of the ADL content. Figure 3 depicts a fragment of the AOM representation of the visual acuity archetype from Figure 2. As it can be observed, there are no explicit, semantic links between the objects obtained with the ADL parser. This last one will return a set of generic objects whose semantics is embedded mainly in string fields. In the figure example, for instance, the kind of a *C_COMPLEX_OBJECT* is shown by means of the *rmTypeName* attribute. It will point out if the object refers to an *ENTRY*, *CLUSTER*, *ELEMENT*, or any reference model class.

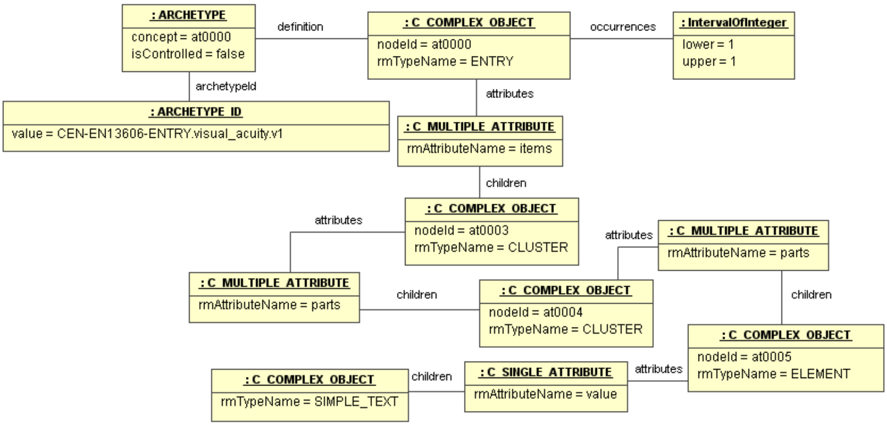

Fig. 3. Extract of the Archetype Object Model of the Visual Acuity archetype.

Thus, the possibilities of reasoning over ADL are currently very limited, as well as the availability of tools to use and manage ADL content is reduced. It does not allow performing any semantic activity on them. However**,** a semantic language would allow.

3. Semantic representation of EHR clinical knowledge

The main purpose of the Semantic Web (Berners-Lee et al., 2001) is to provide a framework in which data can be shared between applications. In order to do this, many technologies have emerged around it with the aim of giving explicit meaning to information, making it easier for machines to automatically process and integrate information. Clinical knowledge, as above stated, is represented by means of archetypes. They are defined using the Archetype Definition Language (ADL). However, this language has some limitations (see section 2.2) that can be solved using a semantic language. Thus, the representation of clinical knowledge, archetypes, as ontologies will be one of our goals and how to carry it out will be explained in this section.

3.1 The need for a semantic representation

The use of ontologies for representing clinical archetypes offers some benefits against the use of ADL. Ontologies allow performing the management of archetypes in an easier and more efficient way. Activities such as comparison, selection, classification and consistency checking can be performed over ontologies in a more generic, easier and efficient way.

In this work we use the Web Ontology Language (OWL), which is the recommendation of the W3C for the exchange of semantic content on the web, for this purpose. In particular, OWL-DL (where DL stands for "Description Logics") is used, because of its decidability and computability nature, offering enough expressiveness and the possibility of reasoning over the information that it describes.

OWL allows making annotations on classes or properties and semantic similarity functions are also available in the Semantic Web community. Thus, these resources help performing all these management related tasks. For instance, the selection of the set of archetypes to be

used in a clinical information system may be supported by semantic similarity functions and by semantic search filters based on the annotations of the archetypes.
Also, another benefit from the use of OWL is related to terminologies. They are very important in the medical field and some of them such as SNOMED-CT (SNOMED-CT) are currently in the process of adapting their representation to Semantic Web environments, so that OWL models for them are under development. Having the representation of both clinical and terminological information in the same formalism would facilitate better clinical knowledge management and would enrich archetypes by adding more information to them.
Moreover, an archetype described in OWL might guarantee the consistency of the knowledge which cannot be granted by ADL (see section 2.2). To grant it, there is the need of implementing additional mechanisms. In addition to this, the access to clinical information described in OWL can be also done in a more natural way. OWL modelling brings all the information concerning a particular term together (code, definition, bindings, translations ...).
Among its benefits, the representation of archetypes in OWL makes the use of tools developed by the Semantic Web community possible. This community has been working for years in methodologies and tools for comparing different ontologies, merging them, identifying inconsistencies and so on. Also, OWL is continually being improved and there is currently a draft version of OWL 2.0 (OWL 2.0). Moreover, different technologies and languages for querying, defining rules and exploiting OWL content are in progress.

3.2 Development of an ontological representation for EHR clinical knowledge

The EN 13606 clinical standard and the OpenEHR specification are based on the dual model-based architecture. So, a first stage in our work was to do a semantic interpretation of clinical archetypes, analyzing their reference and archetype models. Figure 4 illustrates some of the main classes of the ontological representation for the EN 13606 standard case. As it can be observed, concepts from the reference and archetype models are put together expressing the archetype structure in a more comprehensible way. For instance, concepts such as *archetype*, *archetype description*, *archetype description item*, *occurrences*, *cardinality* or *archetype term* exist in the archetype model, but other ones, which are underlined in the Figure, such as *folder*, *composition*, *section*, *entry*, *element* or *clinical datatype* belong to the reference model. Thus, this modelling decision captures the common features of both standards, these are the mentioned archetype model concepts, and allows including the specific concepts that exist in each one. In fact, the OpenEHR ontological representation will be similar to the shown in Figure 4 for EN 13606 except for the concepts as *folder*, *composition*, *section*..., that is, those which belong to the respective reference model.

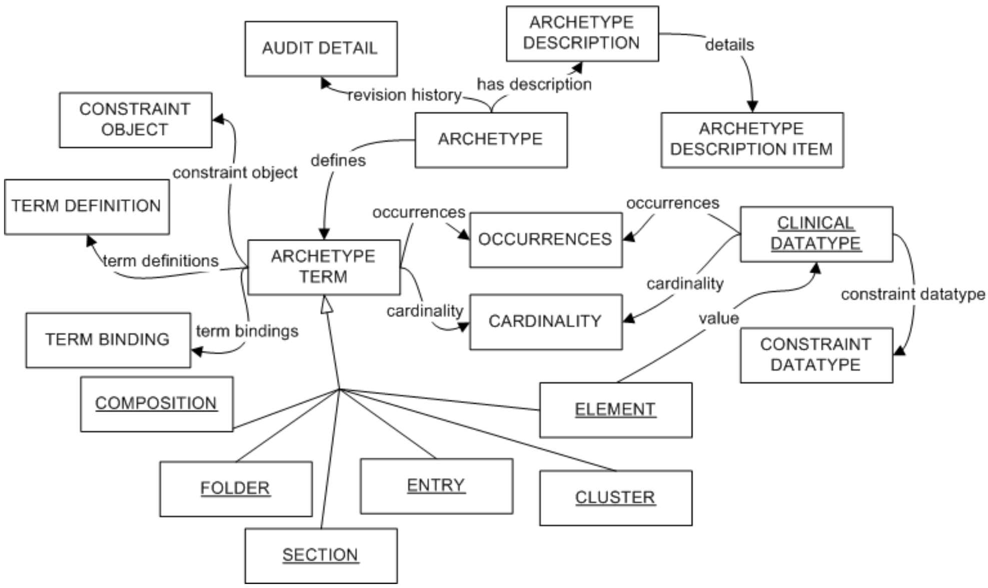

Fig. 4. Fragment of the archetype ontological representation for EN 13606

As a result of the semantic interpretation process (Fernandez-Breis et al., 2006) made for both standards, two main ontologies were built for each one (see Table 2 for details).

- EN13606-SP and OpenEHR-SP: They represent the clinical data structures and data types defined in the reference model of each standard.
- EN13606-AR and OpenEHR-AR: They are the archetype model ontologies; they include some classes of the archetype model, those common to both standards, and import the EN13606-SP and OpenEHR-SP ontologies.

Table 2 gives numeric details of these ontologies in terms of the classes, properties and restrictions. These ontologies allow representing archetypes in a more natural way than ADL does, and all the information regarding to the same clinical term can be accessed together. Moreover, an OWL-based archetype construction approach might guarantee the consistency of the knowledge, which cannot be granted by ADL.

Ontology	Classes	DP	OP	Restrictions
EN13606-SP	68	16	92	227
EN13606-AR	122	76	142	462
OpenEHR-SP	87	14	156	302
OpenEHR-AR	144	75	210	524

Table 2. Details of the OWL ontologies, in terms of classes, data properties (DP), object properties (OP) and restrictions.

3.3 The methodology for obtaining semantic archetypes

This methodology (Martinez-Costa et al., 2009) has been applied to OpenEHR specification and EN 13606 standard and it has been developed using Model Driven Engineering (MDE) techniques. The use of MDE in the development of the methodology allows to take advantage of the tools and experience of the MDE community and to communicate different technical spaces (TS) (Kurtev et al., 2003). The architecture of the solution is depicted in Figure 5, which involves four different technical spaces: Grammar, XML, MDE and Semantic Web. The transformation process is divided in three phases:

i) Representation of syntactic archetypes in MDE from the corresponding representation of archetypes in ADL.
ii) Transformation of models of syntactic archetypes into semantic ones in MDE.
iii) Obtention of OWL semantic archetypes.

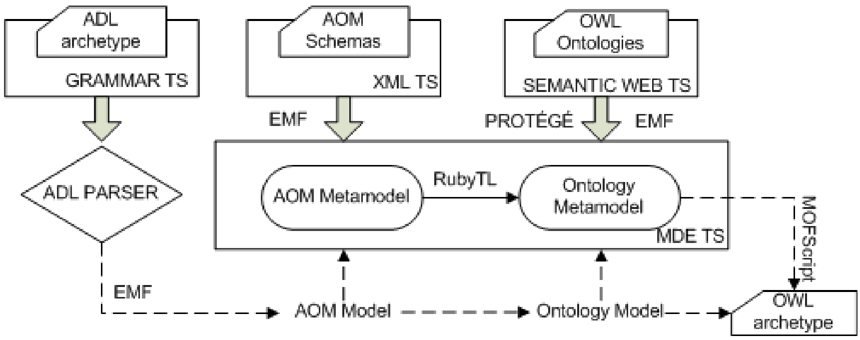

Fig. 5. Architecture of the Methodology for transforming ADL archetypes into OWL

This transformation process has been implemented in a tool which allows for transforming ADL archetypes into OWL for both EN 13606 and OpenEHR standard. This tool is available online (ADL2OWL) and has also been included in the LinkEHR editor (LinkEHR). Next, a more detailed description of the proposed methodology is given and its phases are explained in depth and illustrated through the running example for EN 13606 presented in Figure 2.

3.3.1 Phase I: Representation of syntactic archetypes in MDE

The input to the process is an archetype written according to the ADL Grammar (Grammar TS). This archetype has to be transformed into a generic model according to AOM. This model is generic because its representation is the same for every dual model-based standard. This transformation is carried out by using:

- An ADL parser (ADL-Parser): This is a syntactic parser for ADL, which returns the archetype as a tree of AOM objects.
- An XML serializer (XML-Serializer): This takes an AOM object tree as input and serializes it in XML according to the AOM XML Schema (AOM-Schemas).
- The Eclipse Modelling Framework (EMF): It obtains a metamodel from the AOM XML Schema and allows for serializing the previously obtained XML archetype as a model. Hence, the syntactic representation of archetypes is expressed in MDE.

At the end of this first phase a change of technical space has been produced, from Grammar TS to MDE. Figure 6 shows a fragment of the resulting AOM model for the Visual Acuity archetype example of Figure 2. It will be explained more detailed later in this section.

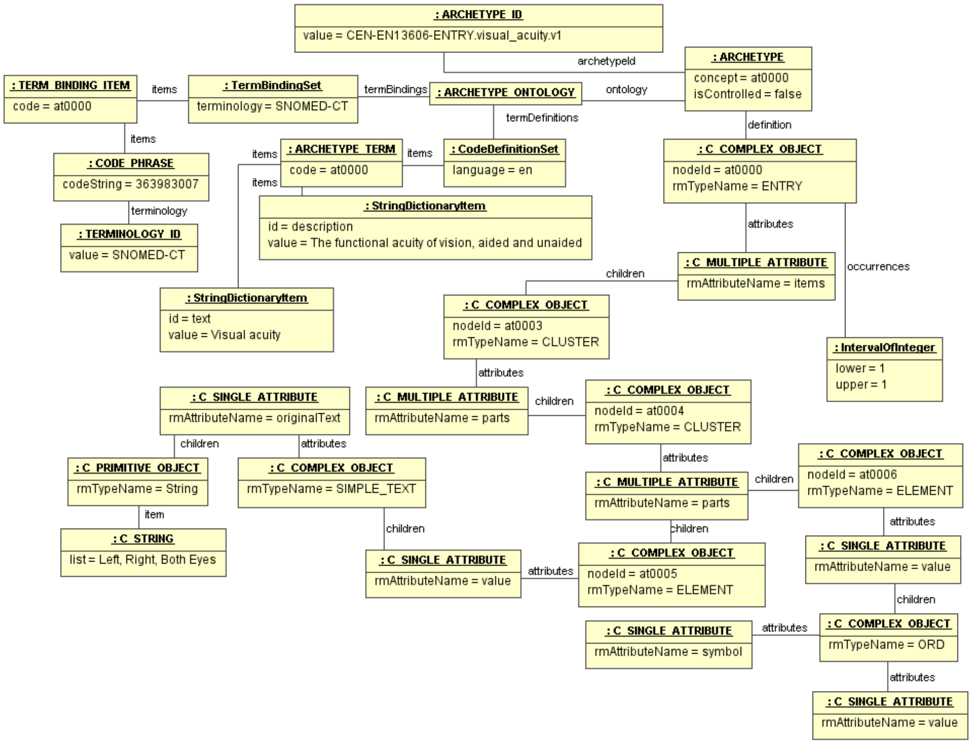

Fig. 6. The AOM model of the archetype obtained in phase I

3.3.2 Phase II: Transformation of models of syntactic archetypes into semantic ones in MDE

The second phase is carried out in the MDE space, and will make use the MDE representation of the archetype models used in both the Grammar and the Semantic Web TS to facilitate the transformation of archetypes from Grammar to Semantic Web. This requires the execution of two tasks:

– MDE representation of the semantic models for the EHR standards. In order to make the transformation, we need to obtain the metamodels for the semantic interpretation of the EHR standards. In this work, metamodels for the CEN-AR and OpenEHR-AR ontologies are obtained as a result of this task. The ODM standard (ODM) defines the representation of OWL ontologies in MDE TS. Protégé (Protege) implements the transformation from OWL to MDE TS and this was the technical solution used to get the metamodels from those ontologies.

– Definition of the transformations between the syntactic and the semantic representation in MDE. This task defines how to obtain a semantic archetype from a syntactic one. A model transformation language, RubyTL (Sanchez-Cuadrado et al., 2006), has been used to define the corresponding set of transformation rules to get semantic archetype models from AOM models.

Figure 7 depicts the corresponding semantic model obtained after performing the previous tasks over the AOM representation of the Visual Acuity archetype. Let us briefly describe some of the correspondences between both metamodels (the AOM metamodel and the semantic one) for the EN 13606 standard. According to the AOM, objects are represented as *C_COMPLEX_OBJECT* and attributes as *C_ATTRIBUTE (C_SINGLE_ATTRIBUTE/C_MULTIPLE_ATTRIBUTE)*. If the definitional part of the visual acuity archetype is analyzed the AOM model will be composed of:

– Nine *C_COMPLEX_OBJECT* having the following values for the pair (*rmTypeName, nodeId*): (1)("ENTRY","at0000"); (2)("CLUSTER", "at0003"); (3)("CLUSTER","at0004"); (4)("ELEMENT", "at0005"); (5)("SIMPLE_TEXT",""); (6)("ELEMENT","at00006"); (7)("ORD","").; (8)("CODED_TEXT",""); (9)("CD","").

– Ten *C_ATTRIBUTES* objects having the value for (*rmAttributeName)*: (1)("items"); (2)("parts"); (3)("parts"); (4)("value"); (5)("originalText"); (6)("value"); (7)("symbol"); (8)("value"); (9)("codedValue"); (10)("displayName").

The generic nature of AOM makes it impossible to make the semantics of these objects explicit, and it is embedded into the string attributes *rmTypeName* and *rmAttributeName*. By analyzing the value of these properties, the following mappings to the corresponding EN 13606 ontology model can be defined:

– Nine *C_COMPLEX_OBJECT* are converted into the following specific elements from EN 13606 reference model: (1)(ENTRY); (2)(CLUSTER); (3)(CLUSTER); (4)(ELEMENT); (5)(SIMPLE_TEXT); (6)(ELEMENT); (7)(ORD); (8)(CODED_TEXT); (9)(CD).

– Ten C_ATTRIBUTE are converted into specific attributes of the previous mentioned types from the reference model: (1)(items); (2)(parts); (3)(parts); (4)(value); (5)(originalText); (6)(value); (7)(symbol); (8)(value); (9)(codedValue); (10)(displayName).

–

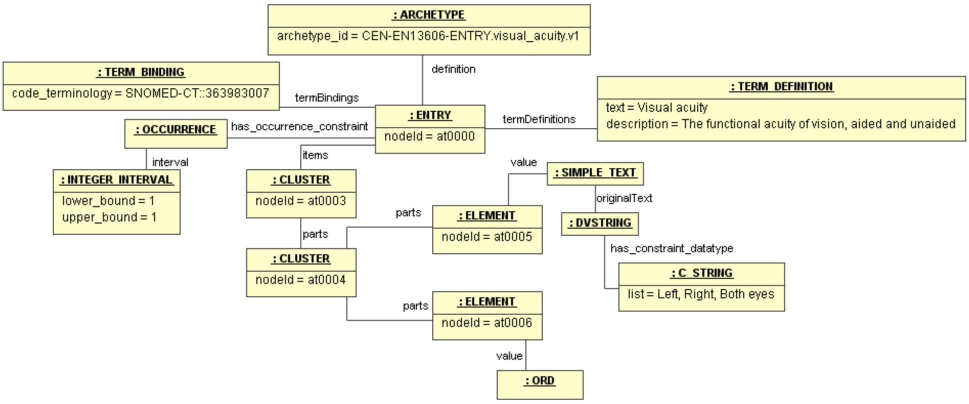

Fig. 7. Ontology model of the archetype obtained in phase II

3.3.3 Phase III: Obtention of OWL semantic archetypes

Finally, archetypes have to be expressed in OWL. In particular, an archetype will be represented as an individual of the class *ARCHETYPE* of the ontology of the particular standard. This transformation implies another technical space change: from MDE TS to Semantic Web TS. For this purpose, the process has to be formalized by specifying the transformation rules that would produce OWL archetypes from the semantic ones. These rules have been written using the model to text transformation language MOFScript (MOFScript). The result of this phase will be the OWL representation of the archetype which fragment is shown in figure 8.

Figure 8 depicts a fragment of the OWL representation of the ADL archetype introduced in Section 2. As it can be observed, an *entry*, which is a subtype of *archetype term*, has several properties as its *code*, *definition*, *occurrences* or *binding* to the SNOMED-CT terminology among others. In contrast, in the ADL archetype representation this information has to be found by string matching of some object attributes. For instance, the definition of this entry in the example should be found by looking up the definition term which code matches with the entry code *(at0000)*. The same situation occurs with its bindings or possible translations.

```
<?xml version = "1.0" encoding="ISO-8859-1"?>
 <rdf:RDF
   xmlns:cen-archetype="http://klt.inf.um.es/~cati/ontologies/CEN-AR-v2.0.owl#"
   ...
  <owl:Ontology rdf:about=""> .. </owl:Ontology>
  ...
  <cen:ENTRY rdf:ID="cen_ENTRY_749d135a-fbc5-4bf5-b761-46ac70728e09">
      <cen:cen_archetype_id rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        at0000
      </cen:cen_archetype_id>
      <cen-archetype:has_occurrence_constraint> ... </cen_archetype:has_occurrence_constraint>
       <cen-archetype:TERM_DEFINITIONrdf:ID="TERM_DEFINITION_b5476f5a">
                <cen-archetype:text rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
              Visual acuity
```

```
            </cen-archetype:text>
            <cen-archetype:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
                The functional acuity of vision, aided and unaided
            </cen-archetype:description>
        </cen-archetype:TERM_DEFINITION>
        <cen-archetype:TERM_BINDING rdf:ID="TERM_BINDING_08b2e630">
           <cen-archetype:code_terminology rdf:datatype="
http://www.w3.org/2001/XMLSchema#string">
                363983007
          </cen-archetype:code_terminology>
         </cen-archetype:TERM_BINDING>
        <cen:cen_items>
            ...
        </cen:cen_items>
   ...
</rdf:RDF>
```

Fig. 8. Extract of the OWL representation of the visual acuity EN 13606 archetype

4. EHR clinical knowledge management

In the previous section, the methodology for obtaining OWL archetypes has been described. The motivation for this was also introduced in that section. There, it was stated that semantic activities could not be efficiently done with ADL but with OWL. Hence, once archetypes are expressed as semantic archetypes in OWL, they can benefit from the Semantic Web technologies. In this section, how such semantic activities can be performed on archetypes is described.

The Archetype Management System (ArchMS) (Fernandez-Breis et al., 2008) has been developed by our research group as the technological solution of the semantic management of archetypes. The objective of this system is to support the execution of clinical, semantics activities over archetypes. ArchMS is built on the idea of a virtual archetype repository for dual-model based EHR standards, whose basic unit is the archetype. In this way, it is capable of working with any dual model-based EHR standard. Given its virtual nature, archetypes are not physically stored in the system but their corresponding URI. Hence, batch processing for ensuring the validity of the links are required.

The current implementation of the system allows for working with both EN13606 and OpenEHR archetypes. It also allows performing two main types of activities with archetypes, namely, classification and search, which are described in the next subsections. Both ADL and OWL archetypes can be input to the system, although the semantic activities are launched on the OWL ones, so the ADL2OWL transformation described in the previous section would be executed for those supplied in ADL. The transformed archetypes are stored in the system. Furthermore, the semantic activities are currently performed over the base of archetypes of the same EHR standard, since the semantic interoperability of EN13606 and OpenEHR archetypes has not been achieved yet.

4.1 Classification of Archetypes

As it has already been mentioned in this chapter, clinical archetypes are specifications of clinical concepts that guide clinical practice and can be considered a template for data acquisition. Hence, the organization of archetypes is a critical issue for optimizing their use, and facilitating their sharing and reuse. Indeed, this would promote the homogeneous clinical practice and facilitate the exchange of clinical information across healthcare institutions.

In this work, the organization of the archetypes is provided by means of annotations, which can be defined in ArchMS with different granularity. The purpose of the annotation is not to facilitate the navigation of the archetype repository for humans but to add semantic metadata to the archetypes, so those can be used to support semantic activities. Hence, semantic annotations are provided. For this purpose, ArchMS makes use of an ontology of annotations, which model how annotations are associated to archetypes. This semantic metadata can be associated to a complete archetype or a term of it. In order to complete the definition process of the annotation, a classification resource is needed. The annotations of the (parts of the) archetype have to be done with respect to an OWL classification resource. Any type of domain ontology can be then used for annotation purposes. In this way, different types of annotations can be done, depending on the type of classification resource. On the one hand, governance ontologies might be used. This would allow for annotating the archetype according to their potential usage and application to particular patients, medical areas, etc. On the other hand, terminological annotations could be used, providing the clinical meaning of the terms of the archetype.

Figure 9 shows how annotations are created. In this example, the running visual acuity example of Figure 2 is associated to the eye concept defined in MESH (MESH), whose code is MESH_A.01.456.505.420, which the code for the eye. This annotation represents that such archetype is related to the eye. The selected archetype is shown on the left, whereas the classifier resource appears on the right. Since both are OWL content, they can be visually represented and browsed as trees, whose root nodes are, respectively, the archetype and the classifier resource. As a result of this annotation, new semantic metadata are added into the system for further exploitation, since the definition included in the archetype is enriched by the semantics associated from the MESH ontology.

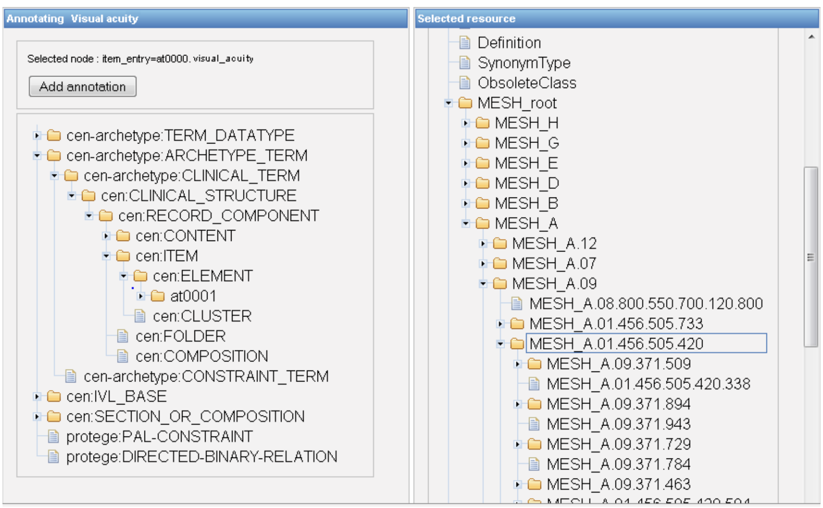

Fig. 9. Annotating the visual acuity archetype in ArchMS

4.2 Search

Semantic annotations are used in ArchMS for facilitating the organization of the archetypes in the repository and for supporting semantic search processes. ArchMS allows for the execution of different types of searches on the virtual base of archetypes, thus exploiting the repository in different dimensions. In general, two main searches can be performed: for similar archetypes, and for archetypes holding some properties, which are described next.

On the one hand, similar archetypes to a given one can be found by doing semantic comparisons in the context of the archetype ontology available for the particular standard. Archetypes are instances of that ontology, so that instances comparison mechanisms can be applied. These mechanisms would take into account the following categories:

- Conceptual proximity: It calculates the ontological distance between the classes in the ontology. For instance, in the context of OpenEHR, two Observation archetypes would be closer and, therefore, more similar than an Observation and a Folder.
- Annotations similarity: The annotations similarity compares the annotations associated to the archetypes. For this purpose, the annotations done with ArchMS are used.
- Property similarity: It compares the sets of properties defined for each archetype, that is, attributes and relations, including the annotations for each property.
- Linguistic proximity: It takes into account external resources to determine semantic distance between the medical concepts defined in the archetype. For this purpose, two external resources are used: Wordnet (Wordnet) and the UMLS metathesaurus (UMLS). Wordnet is more general, whereas UMLS is more exhaustive for medical domains. In order to calculate the similarity, hiperonymy and holonymy are the relations used.

The previous function returns a value between 0 and 1 for every pair of archetypes. Hence, the result of this search is a list of archetypes which is sorted by decreasing similarity. The most similar archetypes will then appear first.

This type of search is the base for suggesting annotations for new archetypes. In this way, the properties of the most similar archetypes can be displayed to the user, which may decide to add such annotations to the new archetype.

On the other hand, users can search for the archetypes that hold some properties. These can be either definitional or annotations properties:

- Definitional properties: They are defined in the proper structure of the archetype, mainly associated to the clinical data types and structures. For instance, we might be looking for archetypes written in a particular language, including an element measured in a certain unit, and so on.
- Annotation properties: They are the annotations associated to the archetype within the ArchMS system. For instance, we might be looking for archetypes related to a particular anatomic part, to a particular disease, for particular types of patients, and so on.

The queries can be mixed, that is, they can include both definitional and annotation properties. The result of this query is the set of archetypes that hold at least one of the requested properties. This set of archetypes is shown sorted by decreasing number of properties held. The archetypes holding more properties will then appear first.

5. Conclusions

In this chapter we have presented an approach for managing EHR-related clinical knowledge from a Semantic Web perspective. This effort constitutes an initial step in the context of the challenging task of achieving semantic interoperability between EHR systems. This would allow health care professionals to manage the complete EHR of patients, independently from which institution generated each clinical session. Semantic interoperability is then an essential factor for improving the quality and safety of patient care, public health, clinical research, and health service management.

To our opinion, the dual model-based architecture that distinguishes two modelling levels, information and knowledge, is the most suitable candidate for that purpose. In this architecture, archetypes represent the knowledge level and are an essential a tool for building clinical consensus in a consistent way and are considered basic to deliver fully interoperable EHRs (Kalra et al., 2008). Archetypes are defined by clinical domain experts using ADL, which is a generic language that does not support the execution of semantic activities. A significant fact of the importance of the dual model architecture and archetypes is the adoption of the European EHR EN 13606 standard by Sweden for their national EHR developments. Its usefulness is also strongly emphasized and its usage recommended by the final report of the Semantic Health project (SemanticHealth Report, 2009).

Hence, in this chapter a representation of archetypes using OWL has been proposed. This required the construction of OWL ontologies for EHR standards such as EN 13606 and OpenEHR standards. For this, the standards were semantically interpreted. A method for transforming ADL archetypes into OWL has also been presented in this chapter, because this allows performing semantic activities such as comparison, selection, classification and consistency checking in a more generic, easier and more efficient way. The OWL technology supports archetype management-related tasks, such as the selection of archetypes to be used in a health information system, the enrichment of archetypes based on the semantics of related-ones, and so on, which are some of the archetype management facilities offered by ArchMS, which has also been presented in this chapter. This system allows for annotating archetypes and performs different types of semantic searches on virtual archetypes repositories.

As further work, we will develop Semantic Web-based mechanisms for transforming OpenEHR archetypes into EN 13606 and vice versa, with the aim of achieving the semantic interoperability between these two dual model standards. The semantic integration of terminologies such as SNOMED-CT and our results should also be researched to enhance the execution of the semantic processes. Finally, we are also developing tools based on the semantic representation of archetypes for supporting the collaborative construction of archetypes and the automatic generation of web data forms.

6. Acknowledgements

This work has been possible thanks to the Spanish Ministry of Science and Innovation through project TSI2007-66575-C02 and to the Spanish Ministry of Industry, Tourism and Commerce through project TSI-020100-2008-577.

7. References

ADL-Parser. http://www.openehr.org/svn/ref_impl_java/trunk/adlparser/
ADL2OWL. http://klt.inf.um.es/~cati/Adl2Owl.htm
AOM-Shemas. http://www.openehr.org/releases/1.0.2/its/XML-schema/index.html

Beale, T. (2001). Archetypes, Constraint-based Domain Models for Future-proof Information Systems. http://www.deepthought.com.au/it/ archetypes/archetypes.pdf

Berners-Lee, T.; Hendler, J. & Lassila, O. (2001). The semantic web, Scientific American, May 2001 pp. 29-37

Brewster, C., O'Hara, K., Fuller, S., Wilks, Y., Franconi, E., Musen, M.A., Ellman, J., Buckinham Shum, S. (2004) Knowledge Representation with Ontologies: The Present and Future. IEEE Intelligent Systems vol 19(1): 72-81

Douglas C. Schmidt. (2006) Guest Editor's Introduction: Model-Driven Engineering, Computer, vol. 39, no. 2, pp. 25-31, Feb. 2006

EMF. http://www.eclipse.org/emf/

Fernandez-Breis, J.; Vivancos-Vicente, P.; Menarguez-Tortosa, M.; Moner, D.; Maldonado, J.; Valencia-Garcia, R. & Miranda-Mena, T. (2006). Using semantic technologies to promote interoperability between electronic healthcare records' information models. *Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS '06*, Aug. 2006 ,pp. 2614-2617, New York, United States

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. (2008). A semantic web-based system for managing clinical archetypes. *Proceedings of IEEE Eng Med Biol Soc 2008,* pp. 1482-1485, Vancouver, Canada

GEHR. http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/index.htm

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition Vol. 5, pp.199-220.

HL7. Health Level Seven – http://www.hl7.org

ISO/TC 215 TR. (2008) Technical report about the EHR, its scope and context http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33397

Kalra, D. & Tapuria, A. (2008). EHR and Clinical Archetypes: time for Clinical Engagement. eHealth Planning and Management Symposium. Copenhagen, Denmark, November 2008

Kurtev, I.; Bezivin, J. & Aksit, M. (2003). Technological spaces: an initial appraisal. *Proceedings of CoopIS, DOA'2002 Federated Conferences, Industrial track*

LinkEHR. http://www.linkehr.com/

Martinez-Costa, C.; Menarguez-Tortosa, M.; Fernandez-Breis, J. & Maldonado, J. (2009) A model-driven approach for representing clinical archetypes for semantic web environments. *Journal of Biomedical Informatics* February 2009, Vol. 42, No. 1, pp. 150-164

Meta-Architecture. http://www.openehr.org/releases/1.0.2/

MESH. http://www.nlm.nih.gov/mesh/

MOFScript. http://www.eclipse.org/gmt/mofscript/

ODM. Ontology metamodel definition specification. http://www.omg.org/cgibin/ doc?ad/2006-05-01.pdf (2006)

OpenEHR. http://www.openehr.org

OWL. http://www.w3.org/tr/owl-ref/

OWL 2.0. http://www.w3.org/2007/OWL/wiki/OWL_Working_Group

Protege. http://protege.stanford.edu/

Sanchez-Cuadrado, J.; Garcia-Molina, J. & Menarguez-Tortosa, M. (2006). Rubytl: A practical, extensible transformation language. *Proceedings of ECMDA-FA 2006.* pp. 158-172

SemanticHealth Report. (2009) *Semantic Interoperability for Better Health and Safer Healthcare,* 2009, Deployment and Research Roadmap for Europe. ISBN-13 : 978-92-79-11139-6

Swedish-decision. A national decision in Sweden on healthcare information standard. http://www.e-recordcompany.eu/attachments/019_Info_about_Swedish _decision.pdf

SNOMED-CT. http://www.snomed.org/

UMLS. http://www.nlm.nih.gov/research/umls

UNE-EN13606. http://www.centc251.org

Van Heijst, G., A. T. Schreiber, & B. J. Wielinga (1997) 'Using explicit ontologies in KBS development'. International Journal of Human-Computer Studies, 45, 183-292.

Wordnet. http://wordnet.princeton.edu/

XML-Serializer. http://www.openehr.org/svn/ref_impl_java/TRUNK/xml-serializer/

Semantic Web

Edited by Gang Wu

ISBN 978-953-7619-54-1
Hard cover, 310 pages
Publisher InTech
Published online 01, January, 2010
Published in print edition January, 2010

Having lived with the World Wide Web for twenty years, surfing the Web becomes a way of our life that cannot be separated. From latest news, photo sharing, social activities, to research collaborations and even commercial activities and government affairs, almost all kinds of information are available and processible via the Web. While people are appreciating the great invention, the father of the Web, Sir Tim Berners-Lee, has started the plan for the next generation of the Web, the Semantic Web. Unlike the Web that was originally designed for reading, the Semantic Web aims at a more intelligent Web severing machines as well as people. The idea behind it is simple: machines can automatically process or “understand” the information, if explicit meanings are given to it. In this way, it facilitates sharing and reuse of data across applications, enterprises, and communities. According to the organisation of the book, the intended readers may come from two groups, i.e. those whose interests include Semantic Web and want to catch on the state-of-the-art research progress in this field; and those who urgently need or just intend to seek help from the Semantic Web. In this sense, readers are not limited to the computer science. Everyone is welcome to find their possible intersection of the Semantic Web.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821