OntoCR: A CEN/ISO-13606 clinical repository based on ontologies

Article *in* Journal of Biomedical Informatics · February 2016

DOI: 10.1016/j.jbi.2016.02.007

4 authors:

Raimundo Lozano-Rubí
Hospital Clínic de Barcelona
14 PUBLICATIONS 27 CITATIONS

Adolfo Muñoz
Instituto de Salud Carlos III
55 PUBLICATIONS 380 CITATIONS

Pablo Serrano-Balazote
Hospital Universitario 12 de Octubre
48 PUBLICATIONS 60 CITATIONS

Xavier Pastor
Hospital Clínic de Barcelona - University of B...
48 PUBLICATIONS 85 CITATIONS

OntoCR: A CEN/ISO-13606 Clinical Repository based on Ontologies

Raimundo Lozano-Rubí[1,2], Adolfo Muñoz Carrero[3], Pablo Serrano Balazote[4], Xavier Pastor[1]

[1] *Medical Informatics. Hospital Clínic, Unit of Medical Informatics. University of Barcelona. Barcelona. Spain*

[2] *Department of Computer Science. Autonomous University of Barcelona. Barcelona. Spain*

[3] *Instituto de Salud Carlos III. Madrid. Spain*

[4] *Hospital Doce de Octubre. Madrid. Spain*

doi:10.1016/j.jbi.2016.02.007

Corresponding Author:

Raimundo Lozano-Rubí, MD

Hospital Clinic

Villarroel 170

08036 Barcelona

Spain

Phone : +34 93 227 92 06

Email: rlozano@clinic.ub.es

ABSTRACT

Objective: To design a new semantically interoperable clinical repository, based on ontologies, conforming to CEN/ISO 13606 standard.

Materials and Methods: The approach followed is to extend OntoCRF, a framework for the development of clinical repositories based on ontologies. The meta-model of OntoCRF has been extended by incorporating an OWL model integrating CEN/ISO 13606, ISO 21090 and SNOMED CT structure.

Results: This approach has demonstrated a complete evaluation cycle involving the creation of the meta-model in OWL format, the creation of a simple test application, and the communication of standardized extracts to another organization.

Discussion: Using a CEN/ISO 13606 based system, an indefinite number of archetypes can be merged (and reused) to build new applications. Our approach, based on the use of ontologies, maintains data storage independent of content specification. With this approach, relational technology can be used for storage, maintaining extensibility capabilities.

Conclusions: The present work demonstrates that it is possible to build a native CEN/ISO 13606 repository for the storage of clinical data. We have demonstrated semantic interoperability of clinical information using CEN/ISO 13606 extracts.

Keywords: Semantic interoperability; electronic health record; clinical repository; CEN/ISO 13606; ontologies; information modeling.

1. **Introduction**

1.1. Objective

With the aim of exploring how to implement a new semantically interoperable Electronic Health Record (EHR) in our institution, we decided to design a clinical repository prototype conforming to CEN/ISO 13606 standard. This repository should be able to automatically incorporate external CEN/ISO 13606 extracts, and to generate CEN/ISO 13606 extracts with the data it contains. It is out of the scope of the system to incorporate or to generate data using other EHR representations than CEN/ISO 13606. This paper describes OntoCR, a proof of concept for a clinical repository driven by ontologies. In this iteration of the repository, we have concentrated on parts 1 and 2 of the standard, leaving aside audit and security issues.

1.2. Background and significance

The deployment of information systems in healthcare facilities has become widespread in recent decades. Nowadays, it is a common form of infrastructure in hospitals, medical offices and diagnosis centers. The main processes at Healthcare facilities are generally well supported, in particular in highly specialized hospitals, where high complexity healthcare delivery and major investment in high tech equipment for diagnosis and therapies requires ubiquitous access, immediacy, concurrency, security and continuous operation (24x7) of all information systems (IS).

However, current systems fail to provide true support to healthcare professionals[1, 2, 3]. New features are often requested, including more structured registration of clinical data, "smart" collection of patient data before clinical contact, intelligent advice for the professional and the patient, offering relevance, pertinence, adequacy, appropriateness and a contextual user interface with interactive capabilities ready to support the clinical decision and the therapeutic intervention in real time at the point of care[4, 5]. Semantic interoperability is an essential factor in achieving benefits from EHR systems to improve the quality and safety of patient care, public health, clinical research, and health service management[6]. In this context, standards are essential for the development and deployment of interoperable eHealth systems, and a considerable amount of effort regarding international harmonization is underway[7]. Generic information models, clinical models, ontologies and terminologies have been identified as required artifacts to achieve semantic interoperability[8], but closer integration between these elements is needed[6, 9, 10]

In recent years, research work has been done on both technical transformation between different approaches proposed[11, 12] and the construction of standardized repositories for secondary use of clinical data[13, 14, 15, 16]. However, there are very few proposals for real implementations of interoperable EHR systems for primary use.

1.3. The CEN/ISO 13606 standard

The overall goal of the CEN/ISO 13606 standard[17, 18, 19, 20] is to define a rigorous and stable information architecture for communicating the EHR (all or part of it) of a single patient. The aim of the standard is to support the interoperability of systems that need to communicate EHR data while preserving the original clinical meaning intended by the author. Further aim is to reflect the confidentiality of that data as intended by the author and patient.

The approach adopted by this standard is named the ‘dual model approach’. It distinguishes a Reference Model (defined in Part 1), used to represent the global characteristics of health record components, and Archetypes (conforming to an Archetype Model, defined in Part 2), which are formal expressions representing a clinical recording scenario.

The Reference Model (13606 RM) comprises a small set of classes that define the generic building blocks to construct EHRs: folder, composition, section, entry, cluster and element.

The Archetype Model (13606 AM) comprises a set of classes to identify, define and describe archetypes, which are prescribed combinations of the building-block classes defined in the Reference Model.

An archetype is a structured and constrained combination of information entities which standardize information collected when describing instances of a given concept, such as blood pressure measurement.

Both the Reference Model and the Archetype Model (13606 AM) are defined in the standard as UML diagrams. The two models are completely independent; there are no common classes between them. Each kind of entity of the Reference Model used in an archetype is specified as the string-value of a property of the Archetype Model.

An archetype interchange format, called Archetype Definition Language (ADL), is proposed in Part 2 of the standard. Although ADL is optional in the standard, the Clinical Information Modeling Initiative (CIMI)[21], an international collaboration dedicated to providing a common format for the representation of semantically interoperable health information, has chosen ADL as a formalism for representing clinical models.

1.4. Proposed Solution

The CEN/ISO 13606 standard does not define an internal architecture or database design of the storage system. Therefore, when implementing this standard over an existing EHR system, a translation layer is needed between the internal structure of the database and the 13606 RM, to transform the data.

One way to avoid complex translations from the internal structure to the Reference Model is to use the latter in the persistence layer. This is the approach followed in OntoCR. Clinical data are stored in the database as instances of archetypes and as 13606 RM constructs: folders, compositions, sections, entries, clusters and elements. In this way, adding data from an extract to the repository and creating extracts from the repository are simple processes.

Previous work done by our group has focused on designing and implementing a framework for the development of clinical data repositories following an ontology-based approach using OWL 1[22]. This approach is called OntoCRF[23]. In OntoCRF, the repository is independent of content specification. All the required information to define a new project is explicitly stated in ontologies. The user interface is built automatically on the fly as web pages, while data are stored in a generic repository. In OntoCRF, data structures are modeled in an ontology according to a specific meta-model that defines the available elements that can be used to build an application, mainly for GUI definition.

The proposed solution is to build OntoCR by extending OntoCRF, thus achieving a native CEN/ISO 13606 clinical repository driven by ontologies.

2. **Materials and methods**

The approach followed in OntoCRF is similar to the dual model of CEN/ISO 13606. In OntoCRF, the information model corresponds with both the database model and the meta-model, which remain unchanged. The knowledge model corresponds with ontologies specifying the content. Using Protégé[39] we have extended the OntoCRF meta-model by incorporating both 13606 RM and 13606 AM, enabling the capability of representing clinical data that conforms to the CEN/ISO 13606 standard. Archetypes in OWL 1 format can be uploaded into the system to build specific applications and specific patient data can be queried as CEN/ISO 13606 extracts.

Figure 1 shows the general architecture envisioned:

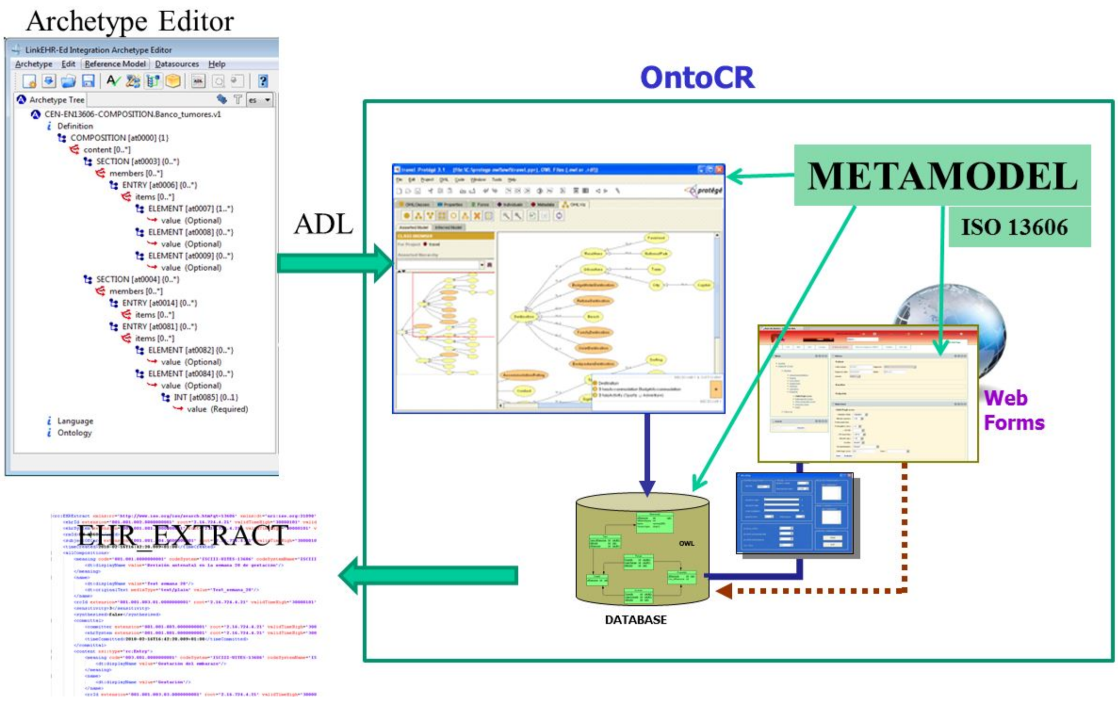


Figure 1 - The general architecture of OntoCR. Archetypes in ADL format are loaded in OntoCR as OWL models. These archetypes are used to build web applications. Instance data can be retrieved as CEN/ISO 13606 extracts.

OntoCR is a native CEN/ISO 13606 system, driven by ontologies, organized in a three-layered model:

1. meta-model layer: meta-model containing definition of CEN/ISO 13606 Reference Model and Archetype model, ISO 21090 data types, SNOMED CT structure, and OntoCRF meta-model for GUI definition;
2. model layer: representation of each detailed archetype and application;
3. instances layer: specific patient data - web forms are automatically generated on the fly to record data specified by archetypes.

2.1. Meta-model description

In OntoCR, we intend that not only communication, but also storage and data capture be available. CEN/ISO 13606 Reference and Archetype Models, ISO 21090 data types, and SNOMED CT structure are modeled and integrated as OWL ontologies, which constitute a meta-model for constructing detailed archetypes.

As a design principle, OWL elements are used when available. This avoids, for example, the need to create a new class list to represent lists of properties, because OWL has its own model to define properties.

Although defined classes in the standards will generally become classes in the meta-model, and defined properties in the standards will generally become properties in the meta-model, the modeling process is not so straightforward. CEN/ISO 13606 and ISO 21090 standards are information models, not conceptual models. To achieve the meta-model, there has to be a transformation process from the information models specified in the standards to the conceptual models represented by ontologies. For example, when a codified item is communicated between two different systems, using de ISO 21090 CD data type, not only the corresponding code is communicated, but also the code system to which it belongs to, the code system version, the code system name, etc. A direct representation of type CD as a class with properties code, codeSystem, codeSystemName, codeSystemVersion, and so on, would result in repeatedly storing data about the code system used, with each code, which would be very inefficient. Representing the relationship between codes and code system conceptually, as explained below, produces a more efficient, consistent and scalable system.

2.2. Modeling data types: ISO 21090

The first element to be modeled is the data type system used.

ISO 21090 standard provides a set of data type definitions for representing and exchanging basic concepts that are commonly encountered in healthcare environments. The different data types are represented as classes, with a number of generalization / specialization relationships among them. For example, a character string is defined in ISO 21090 as:

```
type ST = class (
  validTimeLow : characterstring,
  validTimeHigh : characterstring,
  controlActRoot : characterstring,
  controlActExtension : characterstring,
  nullFlavor : NullFlavor,
  updateMode : UpdateMode,
  flavorId : characterstring,
```

```
    value : characterstring,
    language : characterstring
    translation : SET(ST.NT)
  )
```

In OntCR each one of the classes defined in ISO 21090 has been translated as an OWL class in the ontology, and the generalization / specialization relationships have been maintained. The exceptions are the types defined on a parameter, such as Set(T), where a new class is needed for each combination of the basic type and the parameter type.

Each enumeration, for example NullFlavor, is modeled as a class, and the different values of the enumeration are represented as instances of the corresponding class.

2.3. Reference model: CEN/ISO 13606

Reference model CEN/ISO 13606 is conceptually modeled in an ontology that imports the ISO_21090.owl ontology. In general, each one of the classes defined in the standard is modeled as a class in the ontology, for example RECORD_COMPONENT, COMPOSITION, ENTRY, etc. Each one of the properties defined in the standard is modeled as a property in the ontology. In this way, a single extract of a particular patient becomes an instance in the ontology.

We have previously commented on the design principle for using build-in elements of OWL. Following this principle, defined associations in the RM, such as EHR_EXTRACT.all_compositions, COMPOSITION.content, CONTENT.members, ENTRY.items or ITEM.parts, are implicit in the structure of the ontology. For example, we did not model the attribute ENTRY.items as a generic property with the domain ENTRY and range ITEM. Each specific ENTRY (for example “Tumour data”) has specific OWL properties (for example size) whose range is an ISO 21090 datatype (for example INT)

Properties which are defined in a layer are filled in with values from the lower layer. For example, the definition of an archetype may contain the patient’s age as a property, the value of which will be filled with the specific patient age. A three-layer model implies defining properties in two layers: the model layer (the values are provided in the instances layer) and the meta-model layer (the values are provided in the model layer).

Therefore, properties defined in the standards can be divided into these two categories. Here are examples of each:

- model layer: EHR_EXTRACT.subject_of_care (identifier of the subject of care from whose EHR the EHR Extract was created) and RECORD_COMPONENT.rc_id (the globally-unique identifier by which a node in the EHR hierarchy is referenced). These values depend on each specific instance. These properties are defined as properties of classes.
- Meta-model layer: RECORD_COMPONENT.archetype_id (the identifier of the archetype node) and RECORD_COMPONENT.meaning (the standardised clinical or administrative concept to which the name attribute has been mapped). These values describe the archetype used. These properties are defined as properties of meta-classes.

Properties of instance data are modeled as usual, and a class *RECORD_COMPONENT* has been defined with properties such as *rc_id*. A set of meta-classes has been defined to represent the properties of archetypes and extracts. For example, the meta-class *RECORD_COMPONENT_def* has the properties *archetype_id* and *meaning* among others. The class *RECORD_COMPONENT* is defined as an instance of *RECORD_COMPONENT_def*. The same method has been followed for the rest of elements of the Reference Model, namely FOLDER, COMPOSITION, CONTENT, ENTRY, SECTION, ITEM and CLUSTER. The classes with the corresponding name define the properties of instances and the meta-classes with the suffix "_def" define the properties of archetypes nodes.

A challenging task is to decide which properties belong to each category. The approach followed has been to try to figure out which data will appear with the same value in each instance (for example, the attibutes ehr_system, meaning) and which will appear with different values for each instance (for example, subject_of_care, rc_id). The first group is likely to be represented at the meta-class level , and the second at the class level.

The result is EN_13606_RM.owl ontology shown in Figure 2:

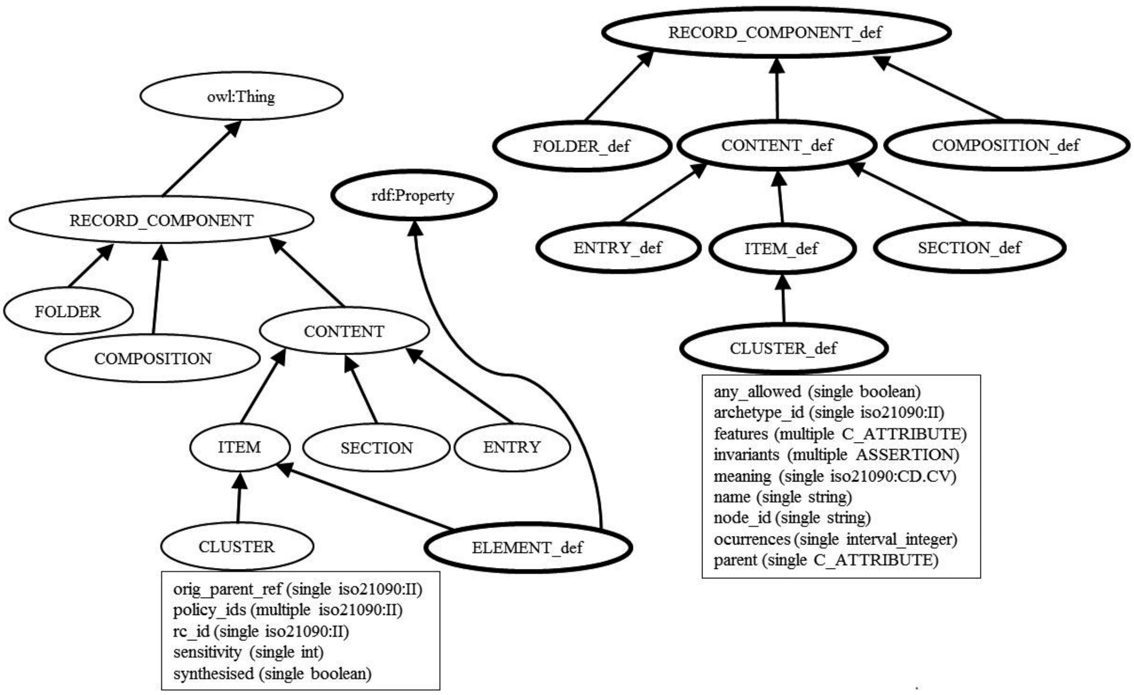


Figure 2 - OWL representation of Reference Model. As an example, the properties of the *CLUSTER* class are shown separated into model and meta-model layers. Ovals with thick border represent meta-classes, the other represent classes. Arrows with closed head represent the relation subClassOf. Rectangles show the properties of CLUSTER_def and CLUSTER. For the sake of readability, the relations of instantiation have been omitted, but Classes RECORD_COMPONENT, FOLDER, COMPOSITION, CONTENT, ITEM, SECTION, CLUSTER and ENTRY are instances of the corresponding classes ending in "_def".

2.4. Archetype model: CEN/ISO 13606

The process followed with the archetype model was similar to the process followed with the reference model. Associations such as C_ATTRIBUTE.features or C_OBJECT.children were modeled implicitly in the structure of the ontology as the RDF properties of the class *C_ATTRIBUTE* and class *OBJECT* respectively.

An archetype represents the recording of a clinical concept that will have multiple specific instances, so all properties of 13606 AM are modeled at the meta-model layer as properties of metaclasses. Their values will subsequently be filled in at the model layer when defining detailed archetypes.

Figure 3 shows the main elements of the ontology:

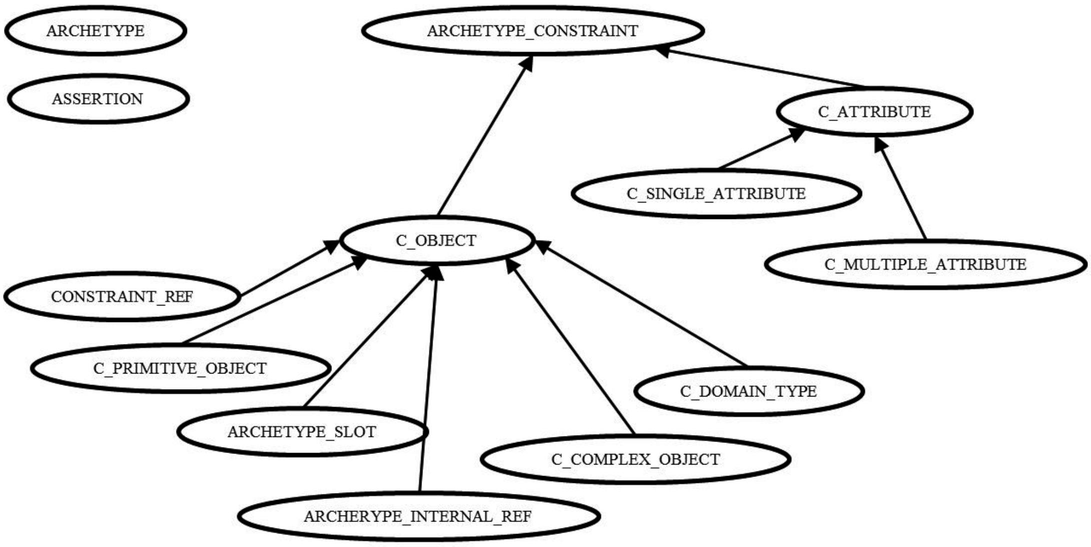


Figure 3 - OWL representation of Archetype Model. Ovals represent meta-classes, arrows represent the relation subClassOf.

2.5. Linking the RM and Archetype model

The link between the RM and AM model is provided in the standard by the property C_OBJECT.rm_type_name:String [1]. We have defined the classes of the RM as subclasses of *C_COMPLEX_OBJECT,* as shown in Figure 4. By doing so, each instance of *C_COMPLEX_OBJECT,* can be modeled as an instance of the corresponding RM class , thus being a stronger way to link both models.

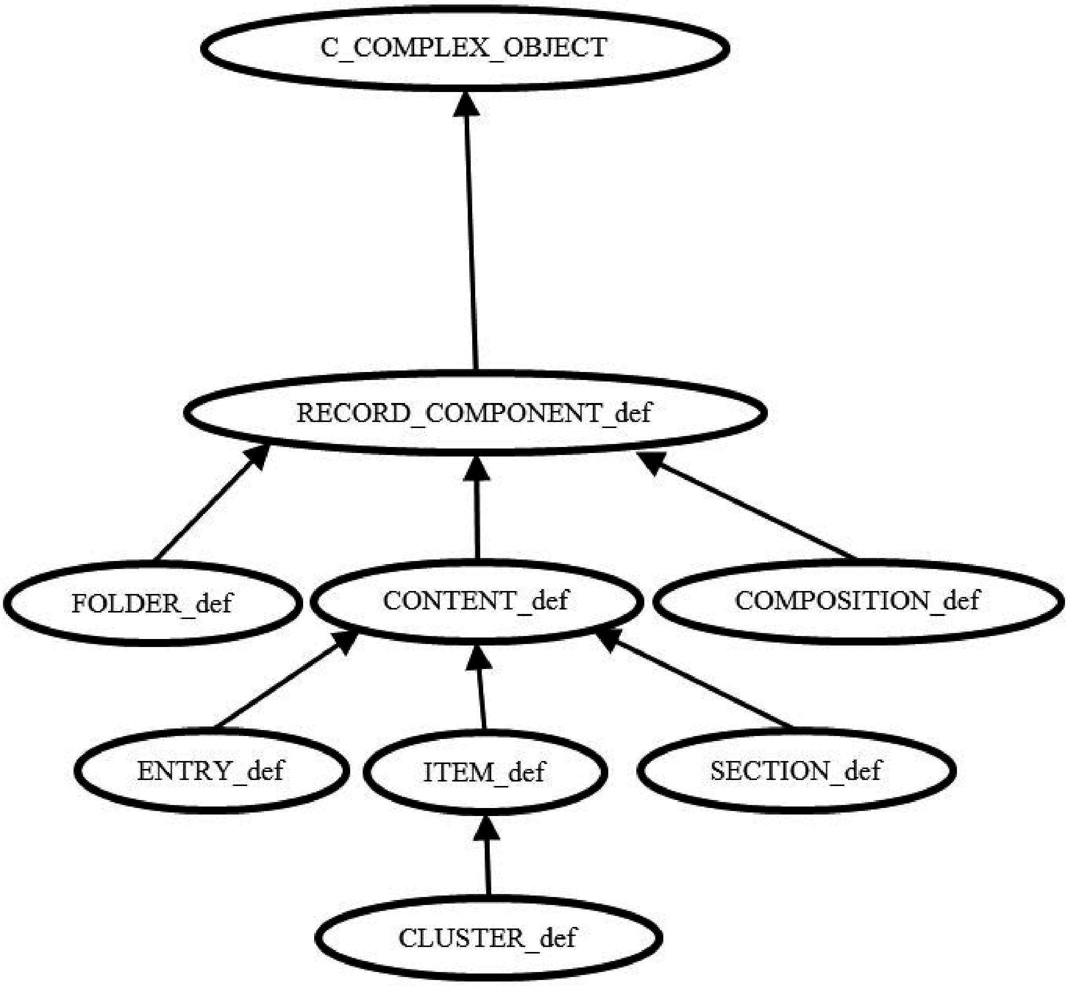


Figure 4 - OWL integration between Reference Model and Archetype model. The classes of the RM become subclasses of C_COMPLEX_OBJECT. Ovals represent meta-classes and arrows the relation subclass of.

2.6. Terminologies

CEN/ISO 13606 not only allows for, but also highly recommends the use of controlled vocabularies. ISO 21090 includes data type CD.CV to represent coded values. This class includes a set of properties to identify the code system to which the code used belongs, but the values of these properties are strings, so the coding system itself is not conceptually represented.

In OntoCR, each vocabulary is represented by a class which is an instance of the meta-class *Codification_system* and is a subclass of the *CD.CV* class. Figure 5 shows an example of SNOMED CT.

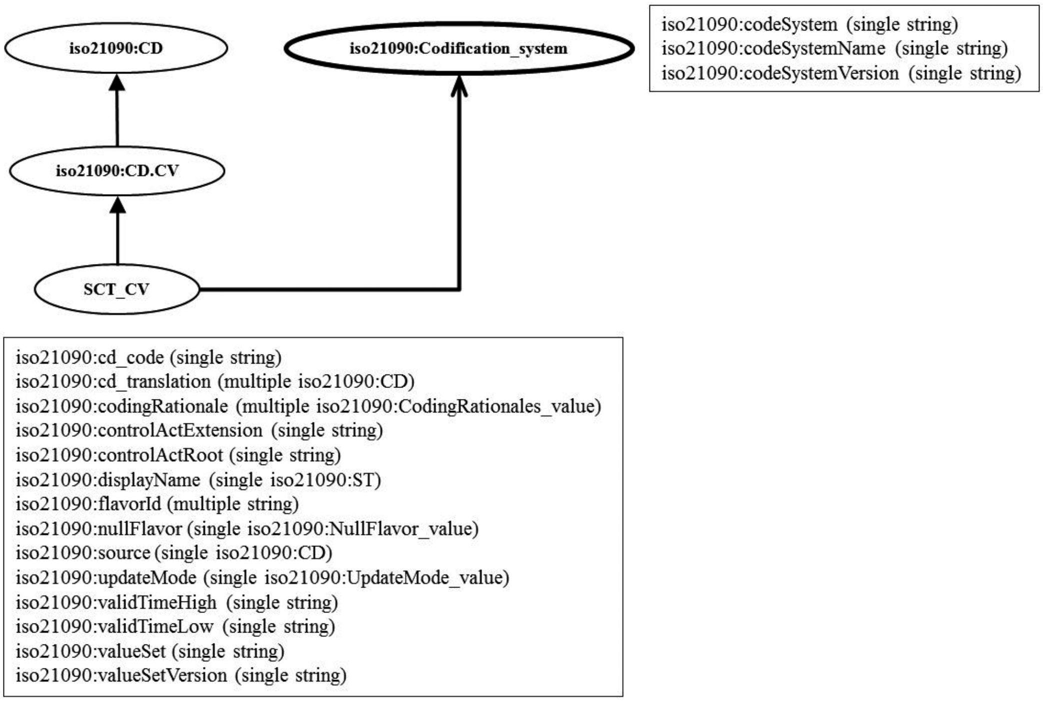


Figure 5 - OWL representation of vocabularies. Ovals with thick border represent meta-classes, the other represent classes. Arrows with closed head represent the relation subClassOf, arrow with opened head represent the relation instantiation. The class SCT_CV is subclass of iso21090:CD.CV and iso21090:CD, and is an instance of the metaclass iso21090:Codification_system. The rectangles represents de properties of classes iso21090:Codification_system and SCT_CV.

This approach allows for the non-repetition of common information for each instance. Information belonging to the vocabulary being used is stored only once. For this proof of concept the required SNOMED CT concepts are predefined in the ontology. In a real scenario this model should be integrated with the use of a terminology server and CTS2 services.

2.7. Detailed archetype representation

CEN/ISO 13606 archetypes can be represented as instances of the meta-model described above. They can be created directly, by defining the archetype from scratch, or by transforming a previously existing ADL archetype into its OWL version. A software tool has been developed to perform such a translation. It produces an OWL version in accordance with the OntoCR meta-model.

3. Results

This approach has demonstrated a complete evaluation cycle involving the creation of the meta-model in OWL format, the creation of a simple test application, and the communication of standardized extracts that, under the umbrella of a research project involving two other organizations, have been sent to a normalized repository in the project EHR platform[24].

The main elements of CEN/ISO 13606 RM and AM have been modeled. As the project is a proof of concept, we have, for the moment, omitted some RM classes relating to audit and attestation information, not needed in our scenario, and some not mandatory properties.

Among the archetypes built to test OntoCR, one of them gathers certain basic information and the clinical stage of breast cancer samples for a tumor bank (Figure 6). Figure 6a partially shows the archetype in ADL format. Figure 6b shows its OWL representation in OntoCR. The clinical stage of the tumor (node at0041 of the archetype) is modeled as an OWL property which is an instance of *ELEMENT_def* . At the same time, this property is an instance of ComboBox, to be represented graphically. Figure 6c shows a web form automatically built from this representation. Once the archetype is uploaded into the system, a simple application accepts specific patient data (see Figure 6c). It is worth noting that data are directly stored as archetype instances in real time. No other conversions are needed.

(a)

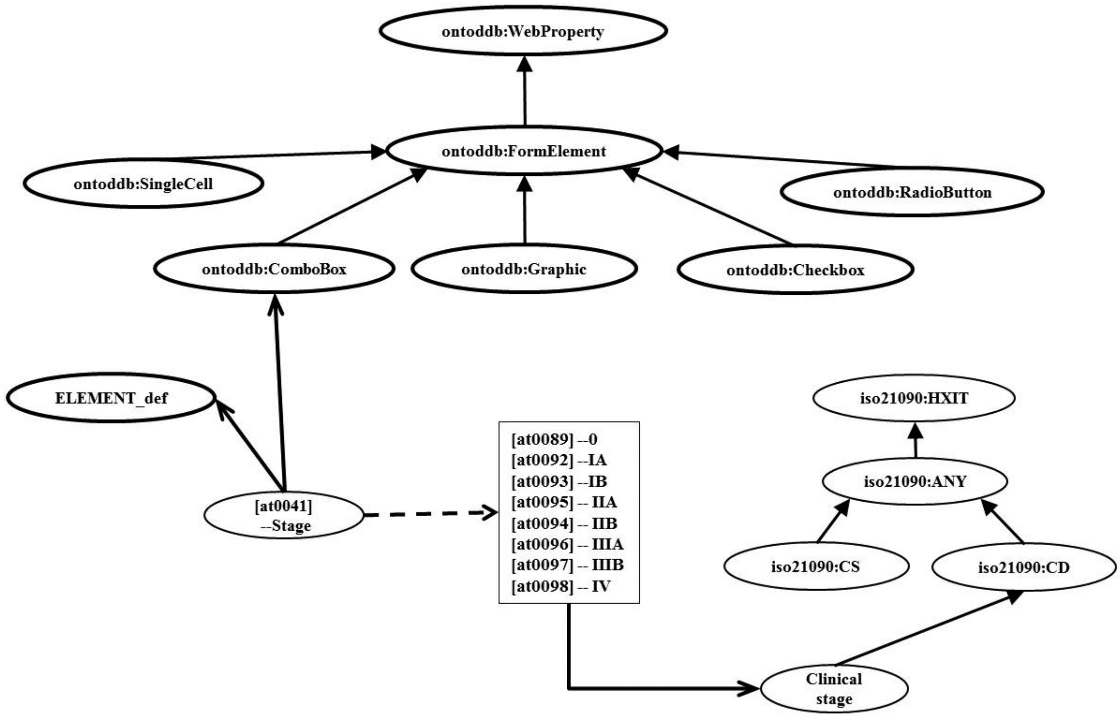
```
archetype
  CEN-EN13606-COMPOSITION.Tumor_Bank.v1

concept
  [at0000]
-----------------------------
-----------------------------

SECTION[at0004] occurrences matches {0..*} matches {  -- TUMOR
  members existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
    ENTRY[at0014] occurrences matches {0..*} matches {  -- Clinical stage
      items existence matches {0..1} cardinality matches {0..*; unordered; unique} matches {
        ELEMENT[at0041] occurrences matches {0..*} matches {  -- Stage
          value existence matches {0..1} matches {
            CODED_TEXT[at0086] occurrences matches {0..1} matches {  --
CODED_TEXT
              codedValue matches {
                CD[at0089] occurrences matches {0..1} matches {  -- 0
                  codeValue existence matches {0..1} matches {*}
                  codingSchemeName  existence matches {0..1} matches {*}
                  displayName  existence matches {0..1} matches {*}
                  qualifiers existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}
                  mappings existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}
                }
                CD[at0092] occurrences matches {0..1} matches {  -- IA
                  codeValue existence matches {0..1} matches {*}
                  codingSchemeName  existence matches {0..1} matches {*}
                  displayName  existence matches {0..1} matches {*}
                  qualifiers existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}
                  mappings existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}
                }
                CD[at0093] occurrences matches {0..1} matches {  -- IB
                  codeValue existence matches {0..1} matches {*}
                  codingSchemeName  existence matches {0..1} matches {*}
                  displayName  existence matches {0..1} matches {*}
                  qualifiers existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}
                  mappings existence matches {0..1} cardinality matches {0..*; unordered;
unique} matches {*}

-----------------------------
-----------------------------
```

(b)



(c)

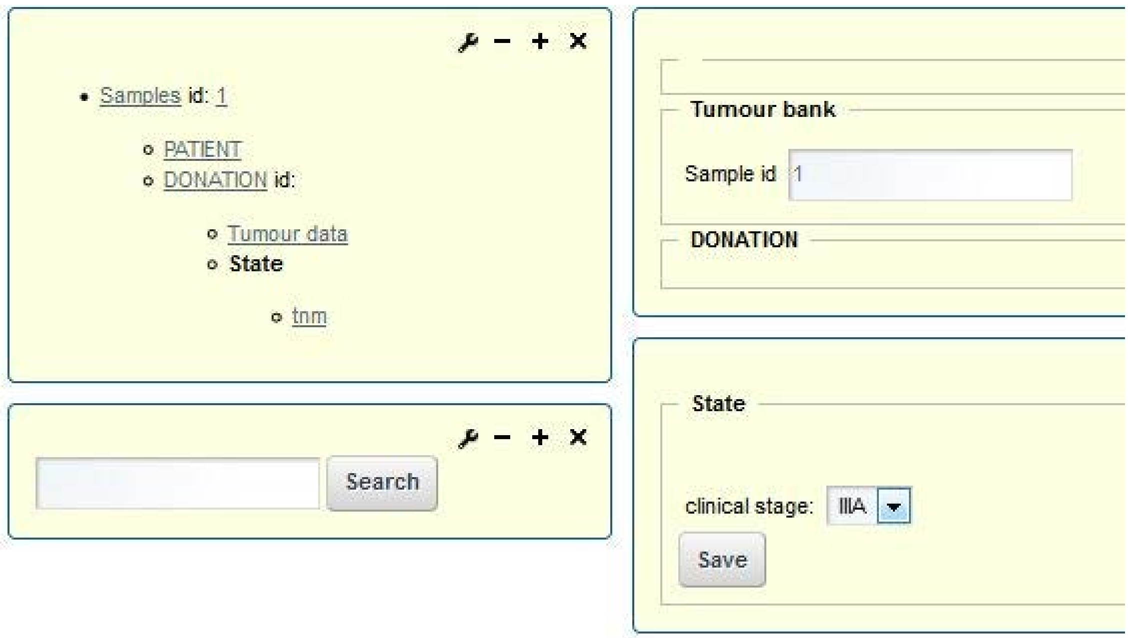


Figure 6.- Representation of the clinical stage of a tumor : (a) in an archetype in ADL format. (b) OWL representation. Ovals with thick border represent meta-classes, the other represent classes. The rectangle represents instances of the class Clinical stage.

Arrows with closed head represent the relation subClassOf, arrow with opened head represent the relation instantiation. The dashed arrow represents the range of the property at0041 (Stage) (c) example of application

In the ADL, the clinical stage is defined as an ELEMENT, node [at0041], whose values are restricted to a value set defined as several CODED_TEXT. In OntoCR, the clinical stage is defined as a property which is, at the same time, an instance of *ontoddb:ComboBox* and *ELEMENT_def*, and which has, as its domain, the *Clinical stage* class. *ComboBox* provides properties to be represented in the GUI. *ELEMENT_def* provides properties to be considered a block of CEN/ISO 13606. The class *Clinical stage* is defined as being a subclass of the class *iso21090:CD* and has, as instances, the different allowed values.

An extraction tool has been developed to produce CEN/ISO 13606 extracts in XML format. For the purpose of validation some fictitious data were introduced and some extracts were sent to the CEN/ISO 13606 repository built by the Instituto de Salud Carlos III[25]. They were correctly validated, successfully uploaded and incorporated to the repository together with extracts coming from other organizations, produced by systems from several providers and created according to different archetypes. The idea is to have a repository of normalized and semantically interoperable information, so new research lines could be started in order to extract new knowledge by means of data mining and machine learning techniques.

The system is, at this moment, at a prototype phase, but so far results are encouraging. Uploading an archetype results in an ontology with a twofold representation:

- a conceptual representation of the clinical concept subject of the archetype. The ontology, seen from this perspective, provides a semantic representation that can be used to retrieve data, and to be linked with other ontologies by describing the domain, etc.
- a representation of the archetype in terms of the reference model capable of accepting patient data as direct instances. This allows for communication of the data as CEN/ISO 13606 extracts.

Populating the archetype only requires the presentation layer, a process that is performed manually by editing the ontology. Because the system is using OntoCRF, both the storage and the user interface are obtained automatically.

4. Discussion

The decision to follow the dual model approach of CEN/ISO 13606 has proven to be a good choice. In OntoCR, an indefinite number of archetypes can be merged (and reused) to build an application. Similar results could be achieved by using openEHR[26]. Our selection of CEN/ISO 13606 is supported by the fact that it is an official and free standard, not subject to the rights of a third-party, and the fact that the Spanish Ministry of Health is in the process of creating CEN/ISO 13606 archetypes to be used at a national level[27].

The use of OntoCRF infrastructure maintains data storage totally independent of content specification. In OntoCR, we use OntoCRF infrastructure to implement the complete information architecture, not only to represent archetypes. This approach ensures greater flexibility and extensibility capabilities, which are necessary requirements of applications in the ever-changing medical field. An alternative is to translate the archetype specifications to their SQL counterparts to record instance data, a process which has several drawbacks. First, an extra step is needed to transform the archetype specification to a relational implementation. Second, this approach leads to a very complex database with a huge number of tables, and is thus difficult to maintain. Third, future modifications of the clinical model may involve modifications to the relational model, which may be difficult to implement in a system with instance data. Fourth, some characteristics of CEN/ISO 13606, such as hierarchies and nested elements, do not fit in well with the relational model and are extremely difficult (not to say impossible) to represent.

The performance of the system when populated in a real scenario remains an open question. Evaluations of OntoCRF[23] showed a linear behavior when uploading and downloading the entire ontology. User interaction with the system in other projects using OntoCRF seems no different than with other systems. Nevertheless, further work needs to be done to evaluate the system in a real scenario.

Although 13606 AM allows for the possibility of coexistence of different types (text or coded value) for an attribute, we feel that to allow for this when designing an archetype would be a poor design choice. When implementing such an attribute in OntoCR, use of a specific type can be forced or, if it is desirable to have both available, two different attributes can be created, one for each type.

We believe the link between 13606 RM and 13606 AM to be very weak. The link is provided by the property C_OBJECT.rm_type_name:String [1]. When instantiated as a node of an archetype, this property should contain the name of the RM type that the node corresponds to, for example “COMPOSITION”, “ENTRY” or “CLUSTER”, but there are no real restrictions as to the value one can use. In the proposed model, there is an ontological link between both models.

ISO 21090 includes the CD.CV data type to represent coded values. This class includes a set of properties to identify the code system to which the code used belongs, but the value of these properties are strings, so the coding system itself is not conceptually represented. In OntoCR, each vocabulary is represented by a class.

The use of ontologies has demonstrated to be a very powerful solution to model the used standards. Extending OntoCRF to create OntoCR has mainly been a conceptual project, involving analyzing the standard and modeling the remaining ontologies as required. In addition, working with ontologies moves the design to the conceptual level, which is more appropriate than technical discussion when modeling clinical concepts.

As mentioned before, current ISO 21090 and CEN/ISO 13606 standards focus on the communication of clinical information and no common semantic layer is assumed. For this reason, the information model is overloaded with elements that actually belong to a conceptual model. As noted in[28], the use of ontologies could solve some major clinical modeling issues, such as whether to put information in the information model or in the terminology model, and how to integrate iso-semantic models. Expressing both the information model and the terminology model in an ontology can help to avoid conceptual overlapping, and thereby facilitate its integration and lead to simpler and clearer archetype design. The proposed system can be enriched as much as needed by integrating ontological representations of other standards, such as CEN/ISO 13940[29], or referencing existing ontological resources[30]. This would be a way of providing a clear ontological commitment for clinical models[8] and formally specifies how information model instances relate to clinical entities. This approach could have a direct consequence: the possibility of simplifying data type and archetype specifications. At the time of writing, OntoCR is not rooted in any upper-level ontology and, thus currently lacks a clear ontological commitment.

Ontologies have been considered promising for decades[31, 32] and now can be envisioned as a solution for common problems in the clinical domain, both as means of heterogeneous data integration[33, 34, 35] and for adding greater cognitive support to applications[36, 37].

Regarding the use of OWL, the expressivity power of the language[38] was adequate to cover the requirements of the standards, and Protégé[39] has proven to be a good tool for ontology editing. The use of OWL allows, to some extent, for the automatic validation of models produced[12]. The addition of reasoning capabilities in the future is a very promising avenue to explore. Reasoning over instance data could provide new knowledge, for example identifying repeated patterns, and the integration with domain ontologies could facilitate clinical checking as drug interactions, drug indication, etc.

The use of OWL has additional advantages. Firstly, it enables users to reuse available knowledge resources[40] and to link them with archetype definitions, thereby adding knowledge to the system. Secondly, there is increasing use of OWL as a formalism for representing clinical models[41, 42, 43, 12, 11]. Moreover, OWL is a standard with wide support among the Semantic Web community[44]. Consequently, tools developed by the Semantic Web community can be directly applied to these models.

Despite all these efforts, there is really no practical experience, as far as we know, of using currently these kinds of models. The most similar work is the proposal of Tao et al[42] to represent the clinical element model (CEM) specification, an information model designed for representing clinical information in EHR systems. They used a three-layer model in OWL: (1) a meta-level ontology that defines the metarepresentation of the CEM; (2) OWL ontologies for representing each individual CEM; and (3) patient data represented as instances of the ontologies on layer 2. The system does not include the possibility to define the graphical user interface. Legaz-García et al[47] propose an archetype management system that uses OWL ontologies to represent CEN/ISO 13606 archetypes. This approach enables the semantic management of archetypes providing interesting functionalities as the transformation of archetypes between standards or their validation. Their proposal does not include how to use the archetypes to record patient data. There are some proposals that use a relational representation. Austin et al[48] developed an EHR server inside a relational database using CEN/ISO 13606 as the information basis for the design of the server. In this case, the choice of a relational representation imposes limits that impede the representation of many features of the standard. Wang et al[49] propose generating relational databases mapping openEHR archetypes to data tables, which implies an extra translation layer. We believe that using a direct relational representation is a less flexible approach and might be difficult to manage versioning.

5. Conclusion and future work

The present work demonstrates that is possible to build a native CEN/ISO 13606 repository for the storage of clinical data. Our approach has been to extend an existing framework for the development of clinical data repositories driven by ontologies: OntoCRF. The similar approach of OntoCRF and CEN/ISO 13606, a dual model separating information and knowledge, establishes a natural way to do it, by adding the conceptual models of CEN/ISO 13606 to the OntoCRF metamodel. Moreover, the proposed system can be enriched as needed by integrating ontological representations of other standards or referencing existing ontological resources.

Furthermore, we have demonstrated semantic interoperability of clinical information using CEN/ISO 13606 between a sender and a receiver, which is the result of independent developments. This is a pioneering experience at an international level.

In relation to future work we are currently working in several directions. At the conceptual level, we are working on the integration of other standards, such as CEN/ISO 13940. In addition, we are considering providing an ontological commitment to OntoCR using an upper-level ontology. The use of BioTopLite2[45] for this purpose may be considered in the future as a potential solution, representing OntoCR classes under “Information object”. This would help to separate the representation of information and what it refers to[46], something usually mixed and confused in current EHR systems.

OntoCR is at this moment a proof of concept focused on design aspects. At the technological level, the next step is to test the system in a real scenario and to evaluate its performance. We are currently transforming the software tools to allow the uploading of archetypes and generation of extracts to web services. We also plan to build an archetype repository.

The possibility of using the repository of normalized information as a source for new knowledge also worth to be studied and will be the subject of future projects.

As a result of these different initiatives, we hope to build a new semantically interoperable EHR based on CEN/ISO 13606 and CEN/ISO 13940.

6. Acknowledgements

This research was partially supported by grants PS09/02076, PI112-01476, PI12-01558, and PI12/01399, “PN 2008-2011- Instituto de Salud Carlos III – Subdirección General de Evaluación y Fomento de la Investigación” of the Spanish Ministry of Science and Innovation, and The European Regional Development Fund.

REFERENCES

[1] Ball MJ, Silva JS, Bierstock S, Douglas JV, Norcio AF, Chakraborty J, et al. Failure to Provide Clinicians Useful IT Systems: Opportunities to Leapfrog Current technologies. Methods Inf Med 47:4-7. 2008

[2] Lehmann CU, Altuwaijri MM, Li YC, Ball MJ, Haux R. Translational Research in Medical Informatics or from Theory to Practice. Methods Inf Med;47:1-3. 2008.

[3] Bouamrane M.M., Tao C., Sarkar I.N. Managing Interoperability and Complexity in Health Systems. Methods Inf Med 54(1), 2015

[4] Shepherd M. Challenges in Health Informatics. Proc 40 HICSS;1-10. 2007.

[5] Mitchell,J.A.; Gerdin,U.; Lindberg,D.A.; Lovis,C.; Martin-Sanchez,F; Miller,R.A.; Shortliffe,E.H.; Leong,T.Y.: 50 Years of Informatics Research on Decision Support: What's Next. Methods Inf Med 50, pp 525-535. 2011

[6] Stroetmann V N et al. Semantic Interoperability for Better Health and Safer Healthcare. Semantic Health Report, 2009

[7] Atalag K et al. Putting Health Record Interoperability Standards to Work. eJHI 5(1) 2010

[8] Martínez-Costa C, et al. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. J Am Med Inform Assoc 2015;0:1–12. doi:10.1093/jamia/ocu013

[9] Semantic Interoperability for Health Network (9). http://www.semantichealthnet.eu/. Accessed January 20, 2016.

[10] Hovenga E.J.; Garde S. Electronic Health Records, Semantic Interoperability and Politics. eJHI 5(1) 2010

[11] Legaz-García M del C, et al. Transformation of standardized clinical models based on OWL technologies: from CEM to OpenEHR archetypes. J Am Med Inform Assoc 2015;0:1–9. doi:10.1093/jamia/ocu027

[12] Menárguez-Tortosa M, Fernández-Breis JT. OWL-based reasoning methods for validating archetypes. J Biomed Inform. 2013;46(2):304–317.

[13] Safran C, Bloomrosen M, Hammond WE, et al., Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association : JAMIA 2007;14(1):1-9. doi:10.1197/jamia.M2273.

[14] Cimino JJ, Ayres EJ. The Clinical Research Data Repository of the US National Institutes of Health. Studies in health technology and informatics 2010;160(Pt 2):1299-1303.

[15] MacKenzie,SL; yatt,MC; Schuff,R; Tenenbaum,JD; Anderson,N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. JAMIA 2012; 19. doi: 10.1136/amiajnl-2011-000508

[16] Pathak J, Bailey KR, Beebe CE, et al. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. J Am Med Inform Assoc 2013;20: e341–e348.

[17] http://www.iso.org/iso/catalogue_detail.htm?csnumber=40784

Accessed January 20, 2016.

[18] http://www.iso.org/iso/catalogue_detail.htm?csnumber=50119

Accessed January 20, 2016.

[19] http://www.iso.org/iso/catalogue_detail.htm?csnumber=50120

Accessed January 20, 2016.

[20] http://www.en13606.org/. Accessed January 20, 2016.

[21] Clinical Information Modeling Initiative (CIMI). http://www.opencimi.org/

Accessed January 20, 2016.

[22] OWL Web Ontology Language Reference. http://www.w3.org/TR/owl-features/ Archived by WebCite® at http://www.webcitation.org/6KDYDWVMT

[23] Lozano-Rubí R, Pastor X, Lozano E. OWLing Clinical Data Repositories with the Ontology Web Language. JMIR Med Inform 2014;2(2):e14
URL: http://medinform.jmir.org/2014/2/e14/
doi:10.2196/medinform.3023

[24] Sánchez de Madariaga R, Cáceres Tello J, Muñoz Carrero A, Moreno Gil O, Velázquez Aza I, Somolinos Cristóbal R. Public electronic Health Record Platform Compliant with the ISO EN13606 Standard as Support to Research Groups. XIII Mediterranean Conference on Medical and biological Engineering and Computing 2013. MEDICON 2013. IFMBE Proceedings. ISBN: 978-3-319-00845-5 (Print) 978-3-319-00846-2 (Online) Sevilla. España 25-28 Sep. 2013.

[25] Muñoz A, Somolinos R, Pascual M, Fragua JA, Gonzalez MA et al. Proof-of-concept Design and Development of an EN13606-based Electronic Health Care Record Service. J Am Med Inform Assoc. 2006; 14(1):118-112.

[26] openEHR. http://www.openehr.org/. Accessed January 20, 2016

[27] Recursos de Modelado Clínico (arquetipos). In spanish. http://www.msssi.gob.es/profesionales/hcdsns/areaRecursosSem/Rec_mod_clinico_arquetipos.htm. Accessed January 20, 2016

[28] Oniki TA, Coyle JF, Parker CG, Huff SM. Lessons learned in detailed clinical modeling at Intermountain Healthcare. J Am Med Inform Assoc. 2014; 21(6), doi:10.1136/amiajnl-2014-002875

[29] http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=58102 Accessed January 20, 2016

[30] Musen,M.A.; Noy,N.F.; Shah,N; et,al. The National Center for Biomedical Ontology. J Am Med Inform Assoc 2012 19:190-195. doi: 10.1136/amiajnl-2011-000523

[31] Chandrasekaran B, Josephson JR. What are ontologies and why do we need them?. IEEE Intelligent Systems, 1999 1:20-26

[32] Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific Am.,May 2001, pp. 34-43

[33] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner, "Ontology-based Integration of Information - A Survey of Existing Approaches," In: Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, 2001, Vol. pp. 108-117.

[34] Ceusters W, Smith B, De-Moor GJ. Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records. IFOMIS Reports 2004. http://ontology.buffalo.edu/medo/CodingAndEHCR.pdf. (Accessed January 20, 2016)

[35] Bettencourt-Silva J, de la Iglesia B, Rayward-Smith V. On Creating a Patient-centric Database from Multiple Hospital Information Systems. Meth Inform Med 2012 51:210-220. doi: 10.3414/ME10-01-0069

[36] Haug PJ, Ferraro JP, Holmen J, Wu K, Mynam K, Ebert M, Dean N, Jones J. An ontology-driven, diagnostic modeling system. J Am Med Inform Assoc. 2013 Jun;20(e1):e102-10. doi: 10.1136/amiajnl-2012-001376

[37] Saleem JJ, Flanagan ME, Wilck N, Demetriades J, Doebbeling B. The next-generation electronic health record: perspectives of key leaders from the US Department

of Veterans Affairs. J Am Med Inform Assoc published online April 18, 2013. doi: 10.1136/amiajnl-2013-001748

[38] Kola J, Wheeldin B, Rector A. Lessons in building OWL Ontology driven applications: OCHWIZ – an Occupational Health Application. Proceedings All Hands 2007

[39] http://protege.stanford.edu/.

Archived by WebCite® at http://www.webcitation.org/6KDYHWgrU

[40] Rector A, Brandt S, Drummond N, Horridge M, Pulestin C, Stevens R. Engineering use cases for modular development of ontologies in OWL. Applied Ontology 2012 7(2) 113-132

[41] Lezcano L, Sicilia MA, Rodriguez-Solano C. Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. J Biomed Inform 2011;44:343–53.

[42] Tao C. et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. Journal of the American Medical Informatics Association 12/2012

[43] Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, Tao C, Parker CG, Haug PJ, Huff SM, Chute CG. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. Journal of Biomedical Informatics 45 (2012) 763–771

[44] W3C Semantic Web Activity Homepage. http://www.w3.org/2001/sw/wiki/OWL Accessed January 20, 2016.

[45] BioTop. http://www2.imbi.uni-freiburg.de/ontology/biotop
Accessed January 20, 2016.

[46] Schulz S, Martínez-Costa C, Karlsson D, Cornet R, Brochhausen M, Rector A. An Ontological Analysis of Reference in Health Record Statements. Conference on Formal Ontology in Information Systems (FOIS 2014), Rio de Janeiro, Brasil, September 22-26, 2014.

[47] Legaz-García MC, Martínez-Costa C, Menárquez-Tortosa M, Fernández-Breis JT. Exploitation of ontologies for the management of clinical archetypes in ArchMS. Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012), KR-MED Series. Graz, Austria, July 21-25, 2012

[48] Austin T, Sun S, Lim YS, Nquyen D, Lea N, Tapuria A, Kalra D. An Electronic Healthcare Record Server Implemented in PostgreSQL. J Healthc Eng. 2015;6(3):325-44.

[49] Wang L, Min L, Wang R, Lu X, Duan H. Archetype relational mapping – a practical openEHR persistence solution. BMC Medical Informatics and Decision Making. 2015 Vol 15. DOI 10.1186/s12911-015-0212-0