MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-338

Genetic Testing Information Standardization in HL7 CDA and ISO13606

Diego Bosca[a], Luis Marco[a], Veronica Burriel[b], Teresa Jaijo[c], Jose M. Millán[c], Ana Levin[b], Oscar Pastor[b], Montserrat Robles[a], Jose Alberto Maldonado[a]

[a] *IBIME Group, ITACA Institute, Universitat Politècnica de València, Spain*
b Research Center on Software Production Methods (PROS), Universitat Politècnica de València, Spain
c Sensorineural Disease Research Group, Instituto de Investigacion Sanitaria IIS-La Fe and CIBERER

Abstract

Communicating genetic testing reports of a patient in a semantically interoperable way remains difficult. Most of the information is stored as non-communicable documents which cannot automatically be processed. The objective of the project was to obtain semantically interoperable genetic testing reports which could be used not only for communication purposes but also for secondary uses, for example clinical trials or clinical decision support. This work describes the first part of the project, the modeling of genetic information reports using EHR standards. We used the Implementation Guide for CDA R2 Genetic Testing Report (GTR) as a basis for modeling the archetypes for both HL7 CDA and CEN/ISO 13606. This approach was validated with the information included in Usher Syndrome reports available at ISS-La Fe. The result of this work were three archetypes following ISO13606 and three archetypes following HL7 CDA model which contained all the information available in both Usher syndrome genetic testing reports and the implementation guide significant parts.

Keywords:

Archetype, HL7 CDA, Implementation Guide, Template, Electronic Health Records.

Introduction

A revolution started in healthcare In the near future healthcare will use both clinical and biological knowledge to provide better personalized treatments. This will allow the delivery of the best treatment based on the patient and disease genetic profiles. Use of genomics in healthcare has the potential to reduce current costs through the use of personalized drugs which controlled their effects on patients and prevented undesirable side effects [1].

In addition, the use of genomics will assist in identifying which genes are responsible for different outcomes of the same treatment on different patients as well as discover diseases patients are susceptible to by identifying mutations that increase their risks.

From a clinical perspective, current efforts are focused on obtaining the complete and personal Electronic Health Records (EHR) regardless of where a patient has been treated. Modern care process requires the participation of multiple clinical actors which need all previously generated and registered information to reduce risks and provide a better care. Obtaining all the information is not an easy task as patients move through different information systems (hospital, primary care, laboratories, etc.) during their care delivery process. The EHR should be semantically interoperable if different organizations need to communicate health records assuming that the data retains the same meaning.

According to the Semantic Health European project [2] there are three basic pillars to EHR semantic interoperability: A standard or conceptual reference model for representing information, shared ontologies and terminologies to define the vocabulary used to describe the data, and high level information structures to define detailed clinical concepts using both information models and terminologies: archetypes. There are several initiatives and standards in healthcare covering aspects, such as ISO13606 [3], openEHR [4], HL7 Clinical Document Architecture (CDA) [5], or SNOMED-CT [6].

However, even if there were standards for this purpose most health information systems are not prepared to handle structured information about the patient's genome. Sending a private report to the clinician who asked for the patient's genetic analysis is currently the standard procedure for most of genetic analysis laboratories. Only the clinician and the patient know the information which is stored in a specific way by the clinician. The reason for this workflow is the nature of the information stored in these reports since they can contain information about a severe disease that the patient may have or develop in the future.

Information currently stored in the reports does not follow any standard protocol. The geneticist chooses the information to be included into the genetic report by creating and completing a template. This approach makes the interoperability of genetic data very difficult.

Some organizations are trying to solve these interoperability problems, and recent efforts by the National Institutes of Health Genetic Testing Registry (NIH GTR) [7] provide a unified way to store genetic tests. Genetic Testing Registry provides a list of field definitions that must, shall or should be included in a genetic test sent to the registry. However, it provides not only a list of valid terms but also the ability to specify new ones. This flexibility hampers interoperability.

To resolve these issues, the archetype model (Archetype Object Model, AOM) provides a way for semantically standardizing genetic testing reports. Archetypes allow the explicit and formal definition of the structure, vocabulary and semantics of domain concepts. These definitions, as part of the knowledge layer, are independent of the health information systems of each hospital in any country.

Work described in this paper shows a way to model semantically the information written in the genetic testing reports by using both the AOM and HL7 CDA standard. The former is the most extended standard to store clinical information as XML documents. Another result of this paper is the definition of the same information (i.e. genetic testing reports) as both ISO13606 archetypes and HL7 CDA archetypes. HL7 CDA archetypes are used as HL7 CDA document templates. In this paper, we will show that using archetypes to represent templates adds some interesting features when compared to current template definition process.

Methods

Archetype Model

Archetype model is part of the ISO 13606 norm [3] and openEHR specifications [4]. Archetypes are domain level concept definitions written in a formal language that provides a reusable and interoperable mechanism to manage the creation, description, validation and querying of EHR. This approach allows the definition of clinical concepts based on a reference model. The reference model lays the foundations for what is valid and invalid when defining archetypes. The overall idea is that clinicians are able to define and interpret archetypes with suitable tools without the aid of technical staff. Archetypes are reusable which means that they can be specialized to better fit the specific requirements of each area. Archetypes are also scalable as they can be combined to create more complex archetypes. Archetypes are useful because they provide a way of binding information structures (reference model) to terminologies and ontologies. The result is the provision of a semantic description of the information.

Since the objective of this work was the inclusion of the genetic testing information into the EHR, we chose HL7 CDA as the reference model to build archetypes, because it is currently one of the most used standards.

HL7 CDA Genetic Testing Report

HL7 CDA [5] defines a model for clinical document persistence based on HL7 Reference Information Model (RIM). A CDA document is a XML file that contains any kind of clinical content in a narrative form and optionally structured into sections or entries like Observation, Procedure, Organizer, Supply and Act.

HL7 Genetic Testing Report (GTR) CDA implementation guide is a document used to specify the structure and contents of genetic testing reports. The 107 page guide developed by clinic genomics and CDA structured documents working groups describes the template (the data set and minimum elements) to store genetic testing reports in a CDA document [8].

Having computable structures that complied with CDA standard (i.e. archetypes defined using HL7 CDA reference model) that are also based on an implementation guide defined by domain experts opens the door to data mining processes applied to the entire EHR. Archetypes provide a new layer to build semantically richer queries to current health information systems. The subject of care transforms from a patient to a complex biological system that includes cellular information, genomic and proteomic expression.

Use case

The work described in this paper was based on the anonymized genetic testing reports from IIS-La Fe, using the archetype model and based on Genetic Testing Report (GTR) CDA implementation guide.

The IIS-La Fe is the Institute of Research in Health from the Hospital La Fe. IIS-La Fe gathers the research activities of all the services from the hospital. These IISs were an initiative of the Spanish Ministry of Economy and Competitiveness. Currently there are 18 certified IISs in Spain.

We selected the reports for the diagnosis and treatment of Usher syndrome patients. Usher syndrome (USH) is an autosomal recessive disease characterized by sensorineural hearing loss and vision loss caused by retinitis pigmentosa and in some cases, vestibular dysfunction. It is clinically and genetically heterogeneous and was the most common underlying cause for deafness and blindness of genetic origin. To date, 10 USH genes were identified as being responsible for this disorder. Furthermore, almost all types of genetic mutations were reported as causative of the disease: nonsense, missense, frame shift, deep intronic, small indels, and splice-site mutations together with large genetic rearrangements involving several exons of these genes. Usher syndrome is an excellent example of a genetically heterogeneous disease with Mendelian inheritance. [9]. Particularly, IIS-La Fe Usher syndrome testing reports contain context (such as demographic information, sample code, report date, family code, or requester information), and genetic report information such as indications, history, sample type, methodology, molecular study, test results, summary and references.

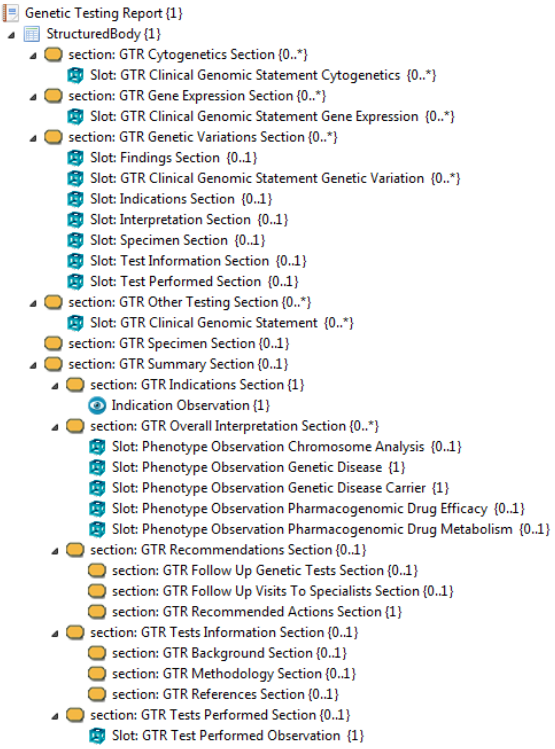
We first had to build the archetypes according to CDA implementation guide. For this purpose LinkEHR [10] was used for modeling the archetypes according to HL7 CDA and ISO13606 models.

For modeling the different archetypes the definitions available for each element in the implementation guide were followed. The implementation guide defined more than 50 different types of entities. It was however unnecessary to model each one in their own archetype. There was only the need to model separately the ones being reused in several places. Combining the entities into a hierarchical tree structure helped in the understanding of the modeled concept.

For each of the relevant entities defined on the guide, available information was extracted for the definition of each of the archetype nodes. The information included its name (*title*), description, type, terminology binding (*code*), occurrences, identifiers (*templateI*and information about which entity types were allowed as descendants in the structure.

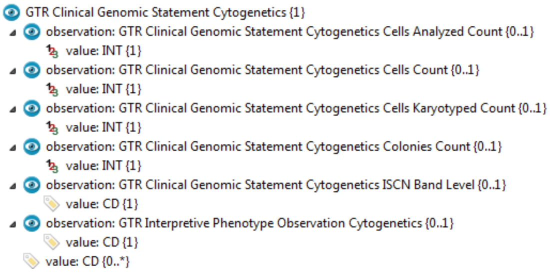
Following this process three archetypes were created, - the archetype describing the general genetic testing report (Figure 1), which included the remaining ones and only defined which sections were included in the document, the archetype for clinical genomic statement cytogenetics (Figure 2), and the archetype for clinical genomic statement genetic variation. These three archetypes contained all the information needed to store all the data available in current Usher syndrome genetic reports.

This design approach assumed that the slots defined in the general archetype for cytogenetics and genetic variation would be filled with the same archetypes in run time. Defining them as separated slots assisted with the evolution of the archetypes.



- Genetic Testing Report {1}
  - StructuredBody {1}
    - section: GTR Cytogenetics Section {0..*}
      - Slot: GTR Clinical Genomic Statement Cytogenetics {0..*}
    - section: GTR Gene Expression Section {0..*}
      - Slot: GTR Clinical Genomic Statement Gene Expression {0..*}
    - section: GTR Genetic Variations Section {0..*}
      - Slot: Findings Section {0..1}
      - Slot: GTR Clinical Genomic Statement Genetic Variation {0..*}
      - Slot: Indications Section {0..1}
      - Slot: Interpretation Section {0..1}
      - Slot: Specimen Section {0..1}
      - Slot: Test Information Section {0..1}
      - Slot: Test Performed Section {0..1}
    - section: GTR Other Testing Section {0..*}
      - Slot: GTR Clinical Genomic Statement {0..*}
    - section: GTR Specimen Section {0..1}
    - section: GTR Summary Section {0..1}
      - section: GTR Indications Section {1}
        - Indication Observation {1}
      - section: GTR Overall Interpretation Section {0..*}
        - Slot: Phenotype Observation Chromosome Analysis {0..1}
        - Slot: Phenotype Observation Genetic Disease {1}
        - Slot: Phenotype Observation Genetic Disease Carrier {1}
        - Slot: Phenotype Observation Pharmacogenomic Drug Efficacy {0..1}
        - Slot: Phenotype Observation Pharmacogenomic Drug Metabolism {0..1}
      - section: GTR Recommendations Section {0..1}
        - section: GTR Follow Up Genetic Tests Section {0..1}
        - section: GTR Follow Up Visits To Specialists Section {0..1}
        - section: GTR Recommended Actions Section {1}
      - section: GTR Tests Information Section {0..1}
        - section: GTR Background Section {0..1}
        - section: GTR Methodology Section {0..1}
        - section: GTR References Section {0..1}
      - section: GTR Tests Performed Section {0..1}
        - Slot: GTR Test Performed Observation {1}

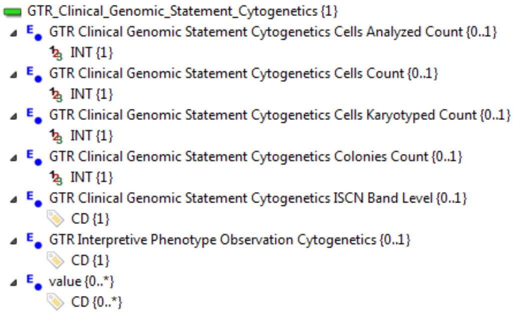
Figure 1 - General genetic testing report HL7 CDA archetype



- GTR Clinical Genomic Statement Cytogenetics {1}
  - observation: GTR Clinical Genomic Statement Cytogenetics Cells Analyzed Count {0..1}
    - value: INT {1}
  - observation: GTR Clinical Genomic Statement Cytogenetics Cells Count {0..1}
    - value: INT {1}
  - observation: GTR Clinical Genomic Statement Cytogenetics Cells Karyotyped Count {0..1}
    - value: INT {1}
  - observation: GTR Clinical Genomic Statement Cytogenetics Colonies Count {0..1}
    - value: INT {1}
  - observation: GTR Clinical Genomic Statement Cytogenetics ISCN Band Level {0..1}
    - value: CD {1}
  - observation: GTR Interpretive Phenotype Observation Cytogenetics {0..1}
    - value: CD {1}
  - value: CD {0..*}

Figure 2 - Cytogenetics HL7 CDA archetype

Once the HL7 CDA archetypes were defined these same concepts were translated into ISO13606. We chose to build HL7 CDA archetypes first because ISO13606 had more generic entities and made the translation process easier (e.g. in general all different kinds of entries in CDA were translated into 'ENTRY' class in ISO13606). Part 3 of ISO13606 standard [11] also states which codes and structures should be used when transforming CDA archetypes into ISO13606. The main problem with this translation was that HL7 has a mechanism to conceptually relate entries that ISO13606 does not (ISO13606 allow the linking of entries at instance level, not at archetype definition level). This was solved by changing the type of CDA entries to other compatible entities in 13606. There were two different ways we could have followed: Changing root class to a section and keeping all the related observations as entries or keeping the root class as an entry and changing the existing related classes to something compatible. In this case, as CDA observations were quite simple, we had to model them as Element classes. An example of the cytogenetics ISO13606 archetype is found in Figure 3.



- GTR_Clinical_Genomic_Statement_Cytogenetics {1}
  - E GTR Clinical Genomic Statement Cytogenetics Cells Analyzed Count {0..1}
    - INT {1}
  - E GTR Clinical Genomic Statement Cytogenetics Cells Count {0..1}
    - INT {1}
  - E GTR Clinical Genomic Statement Cytogenetics Cells Karyotyped Count {0..1}
    - INT {1}
  - E GTR Clinical Genomic Statement Cytogenetics Colonies Count {0..1}
    - INT {1}
  - E GTR Clinical Genomic Statement Cytogenetics ISCN Band Level {0..1}
    - CD {1}
  - E GTR Interpretive Phenotype Observation Cytogenetics {0..1}
    - CD {1}
  - E value {0..*}
    - CD {0..*}

Figure 3 – Cytogenetics ISO13606 archetype

Archetypes provided some interesting features over the implementation guides, such as being able to be processed, the ability to specialize archetypes and easy internationalization of the concepts (a key issue in Spain, with four co-official languages in addition to Spanish).

As a final step we checked that all the information available on current Usher syndrome reports could be represented with this approach. One of the advantages of using a dual model was that archetypes implicitly included all the structures defined in the reference model. This was the case with information included in the Usher syndrome reports such as demographic information or information about the clinician. This information came from the reference model and thus was not explicitly included into the archetypes. Taking all this into account we checked that all the information contained in the reports could be represented with the archetype and the underlying HL7 CDA reference model. This held true even if the structure of the Usher syndrome testing report and the archetype differed (e.g. methodology part on the Usher syndrome report was included inside the test information summary on the archetype). There were some other values that did not have a direct link at first sight, such as the family identifier, very important when dealing with genetic diseases. In this case, we stored them as additional identifiers of the CDA document.

Results

In total, six archetypes were generated for representing current genetic testing reports for Usher syndrome. As seen in Figure 1, the general genetic testing report archetype had several open slots. These slots allowed us to reuse archetypes. For this project, we chose to model only the information already available on the Usher syndrome genetic testing reports. Remaining archetypes (the ones defined in the general genetic testing report archetype) will be defined when we try to model other genetic disease reports.

All the constraints stated in the implementation guide could be defined into the HL7 CDA archetypes. It is worth mentioning that even some complex constraints stated in the implementation guide (e.g. 'this object should not have any other attribute than id') could be easily included into the archetypes. Even if any of the constraints on the guide could not be directly put into the archetypes, archetypes could include constraints as formal definition of assertions over data.

Discussion

Defined archetypes follow HL7 CDA standard and include all the sections needed to correctly define genetic reports structure as defined in the HL7 GTR implementation guide.

We could have chosen other genetic testing report definitions to model the archetypes (such as [6]). However, no such alternatives existed at the beginning of the project. Using the HL7 implementation guide eased the process as the transformation was more direct. If any other definition of genetic testing report was chosen as a basis, an additional step would be needed to decide which classes of HL7 CDA reference model to use for each of the type definition. Creating archetypes for the NIH GTR could be an interesting exercise to aid in the semantic interoperability of genetic data stored on this kind of repositories.

Archetypes also ease lifecycle and evolution of concepts, which is even more important on a changing field like genetic testing. Changes on how we understand the genome or what information is included on genetic reports can be managed by versioning, specializing, or marking some concepts as obsolete.

Specialization also provides a mechanism for the reusability of the archetypes. In this case, not only a general genetic testing report archetype was defined but also a specialized Usher syndrome report. This means that Usher syndrome testing report follows both the restrictions stated in the Usher syndrome archetype and the ones defined in the genetic testing report archetype.

Resulting archetypes provide a formal unambiguous definition of the reports which can be interpreted by clinical and technical staff. They can also be used as structures to map information from several data sources [12] with the objective of automatic generation of CDA and ISO13606 genetic testing reports. This process will make the IIS-La Fe repository semantically interoperable with other hospitals and research centers. This would also provide a way for the inclusion of genetic information into the hospital health information system providing clinicians with genetic testing analysis and conclusions, which would be a step towards personalized medicine and transforming the patient into a complex biological system.

Conclusion

As medicine moves towards personalized medicine, health information systems need to be prepared to store, manage and process this new data. Personalized medicine aims to improve prevention, diagnosis and treatments of patients by taking into account their genetic profile. To assist in this process, archetypes will be allowed to include the genomic information to EHR in a scalable and noninvasive way. This will ease the management and data mining of clinical data, and as a result become a useful tool for clinicians.

To date, even if there are archetypes to deal with genetic-based problems (such as cystic fibrosis review [13] or cancer archetypes available in the Clinical Knowledge Manager [14]), they are generated to deal with medical check-ups and examinations and as a result they leave out related genetic information. Archetypes developed in this project took into account the genetic cause of the problem, and thus provided the clinicians key information for choosing the right treatment. New archetypes could be created to deal with different aspects of the information (e.g. we do not use the same information in a genetic testing report aimed for the patient, for the clinician, or for clinical research).

Combining HL7 CDA standards with the archetype approach has proven effective for modeling genomic information. However, we had some difficulties during the development of the project mostly due to the misalignment of current genetic test reports and the implementation guide.

Archetypes defined give a formal definition of the concepts used in genetic medicine. In the second stage of the project we were able to link the genetic information to the EHR and communicate the Usher syndrome information.

This work also proved that the combination of archetype methodology and HL7 CDA reference model was not only possible but was applied in the modeling of information (genetic information in our use case). Some other experiences, such as [15], also proved that this approach could be applied to other domains.

During the later stages of the project we planned to use created archetypes as mapping templates to standardize genomic information on IIS-La Fe.

Acknowledgments

This work is partially supported by the Spanish Ministry of Science and Innovation under Grant TIN2010-21388-C02-01 and by the Universitat Politècnica de Valencia under Grant PAID-05-11.

It is also important to highlight the support of Cátedra of Tecnologías para la Salud financed by Indra Systems in the development of this work.

References

[1] Armstrong, K. Can Genomics Bend the Cost Curve? JAMA. 2012; 307(10):1031-1032.

[2] Semantic Interoperability for Better Health and Safer Healthcare. Deployment and Research Roadmap for Europe. European Communities, 2009. Available from: http://ec.europa.eu/information_society/activities/health/docs/publications/2009/2009semantic-health-report.pdf.

[3] ISO 13606-2:2008. Health informatics -- Electronic health record communication -- Part 2: Archetype interchange specification

[4] Beale, T. Archetypes: Constraint-based Domain Models for Future-proof Information Systems. Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. Boston (2002), pp. 16-32.

[5] ISO/HL7 27932:2009. Data Exchange Standards -- HL7 Clinical Document Architecture, Release 2

[6] SNOMED Clinical Terms Technical Specification. College of American Pathologists, Northfield, IL, 2001. Available from: http://www.snomed.org

[7] Wendy S. Rubinstein; Donna R. Maglott; Jennifer M. Lee; Brandi L. Kattman; Adriana J. Malheiro; Michael Ovetsky; Vichet Hem; Viatcheslav Gorelenkov; Guangfeng Song; Craig Wallin; Nora Husain; Shanmuga

Chitipiralla; Kenneth S. Katz; Douglas Hoffman; Wonhee Jang; Mark Johnson; Fedor Karmanov; Alexander Ukrainchik; Mikhail Denisenko; Cathy Fomous; Kathy Hudson; James M. Ostell. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. Nucleic Acids Research 2013; 41 (D1): D925-D935

[8] CDA Implementation Guide for Genetic Testing Report (GTR) (September 2011 Draft) http://www.hl7.org/documentcenter

[9] Millán J.M., Aller E, Jaijo T, Blanco-Kelly F, Gimenez-Pardo A, Ayuso C. An update on the genetics of usher syndrome. J Ophthalmol. 2011;2011:417217. Epub 2010 Dec 23.

[10] Maldonado J.A., Moner D., Boscá D., Fernández J.T., Angulo C.,Robles M. LinkEHR-Ed: A multireference model archetype editor based on formal semantics. International Journal of Medical Informatics 78(8)(2009), pp. 559-570.

[11] ISO 13606-3:2009. Health informatics -- Electronic health record communication -- Part 3: Reference archetypes and term lists

[12] Maldonado J.A., Costa C., Moner D., Menánguez M., Boscá D., Miñarro J.A., Fernández J.T., Robles M. Using the ResearchEHR platform to facilitate the practical application of the EHR standards. Journal of Biomedical Informatics 46(4), pp. 746-762, 2012.

[13] Corrigan D. Towards use of OpenEHR Archetypes to support views of Cystic Fibrosis Review Records. 15th Annual Health Informatics Society of Ireland Conference, 18 November 2010, Dublin

[14] Garde S, Chen R, Leslie H, Beale T, McNicoll I, Heard S. Archetype-based Knowledge Management for Semantic Interoperability of Electronic Health Records. Medical Informatics Europe 2009; 2009 September; Sarajevo, Bosnia & Herzegovina.

[15] Moner D. Moreno A., Maldonado J.A., Robles M., Parra C.. Using Archetypes for Defining CDA Templates. Quality of life through quality of information, pp. 53-57, IOS Press BV, Amsterdam. ISBN 978-1-61499-100-7

Address for correspondence

Diego Boscá Tomás, diebosto@upv.es