An Ontology-Based Architecture to Support Language Variants of Model-Driven Electronic Health Records

John CHELSOM[a,b], Stephanie CABRERA[a,b], Zikai HAO[b], Naveed DOGAR[a,c] and Ilyas ADEN[d]
[a] *Fordham University, USA*
[b] *University of Victoria, Canada*
[c] *University of Oxford, UK*
[d] *City University of London, UK*

Abstract. This paper describes the practical implementation of support for multiple language variants of the clinical information models used to define a model-driven EHR system. Beyond that, it describes how a complete EHR system can be created with multiple language variants, using the example of an EHR used for clinical management of patients in a Fracture Liaison Service. A clinical information model was developed in the English language and represented as an ontology. This model was then translated into Spanish and Mandarin, with the experience gained then used to refine the modeling tools. A workshop was then held, where participants used the EHR tools to create language variants of the FLS EHR in German, French, Portuguese, Arabic, Farsi, Urdu and Somali. The results from the workshop are presented here, with a brief summary of the lessons learned .

Keywords. EHR, ISO 13606, HL7 CDA, Internationalization, Language Variant.

1. Introduction

Model-driven Electronic Health Record (EHR) systems use an underlying clinical information model to define the structure of the health record, to govern the interaction with clinical system users as they enter new information or interact with existing information, and to define the interfaces with external systems as clinical information flows in and out of the EHR. The most prevalent architecture for model-driven EHR is defined by the ISO 13606 standard, which specifies (among other things) a base Reference Model for the structure of electronic health records and an Archetype Definition Language, which can be used to define the models (called archetypes) that are based on the Reference Model and drive the operation of specific EHR systems [1].

A model-driven EHR generally maintains a proper Separation of Concerns [2] between the information structure of the health record and the user interface for the entry of new information or the viewing of existing information. This separation of concerns facilitates the implementation of multilingual EHR systems, where clinical information may be created or viewed by users who speak a variety of different languages. The ISO 13606 standard, and the ongoing activities of the openEHR Foundation [3], incorporate methods for supporting multiple languages in the definition

of clinical information models using the Archetype Definition Language. Such models can be created using the openEHR Clinical Knowledge Manager, as demonstrated in a number of studies on the sharing of Archetypes in different languages [4, 5].

This paper describes an alternative implementation of a model-driven EHR system which can support multiple language variants of the same clinical information model. Development was driven by requirements to support three main Use Cases:

1. Using a single information model for gathering data for clinical studies, where the study is made across territories, which have different native languages.
2. Health records gathered and viewed in territories which have more than one official language, where there may be a legal requirement to support two or more languages as the health record is created.
3. Health records gathered in territories where the official language used in healthcare provider organizations may be different from the native language spoken by some patients; here there may be a requirement to allow patients to access their record in a language different to the one in which it was gathered.

The remaining sections describe the methods used to support multilingual health records in the open source cityEHR system [6], with a brief description of the results of implementing an information model for an EHR in a Fracture Liaison Service using a base language of English and variants in Spanish and Mandarin. We present the lessons learned from this implementation and its extension to additional language variants.

2. Method

cityEHR [6] is a model-driven EHR that uses the ISO 13606 standard for definition of Clinical Information Models and the HL7 Clinical Document Architecture (CDA) [7] for storage of information in the electronic record. Instead of the Archetype Definition Language, cityEHR uses an Information Architecture based on the ISO 13606 Reference Model, but represented as an ontology, using the OWL/XML standard [8].

Clinical Information Models are created as OWL/XML ontologies from the building blocks in the base ontology architecture, using spreadsheets to guide clinical analysts in developing the models. The information models are then transformed, using XSLT, to a set of forms for clinical data input, using the XForms standard, with associated HL7 CDA documents, that are used for data input and for storing information in the longitudinal patient record; this process is shown in Figure 1.

The overall cityEHR system acts as a framework in which multiple EHR systems can operate, each defined by their own Clinical Information Model. A single model can be used to support local installations, including installations in different international settings. Localization is achieved by using Clinical Information Directories of ISO 13606 Entries or Elements, linked to the information model, but customized for each local deployment. Examples of such directories include lists of clinics or primary care practices, drug lists, and 'boilerplate' text for clinical correspondence (letters) or reports.

Internationalization is achieved by combining the methods for localization, with support for language variants of the three main sources of information required in a fully functional, model-driven EHR.

1. The Clinical Information Model (as an OWL/XML ontology)
2. Clinical Information Directories (as collections of ISO 13606 Entries or Elements, held as snippets of HL7 CDA)
3. System and application configuration (labels for menu items, messages and other text in the user interface, represented in a custom XML vocabulary)

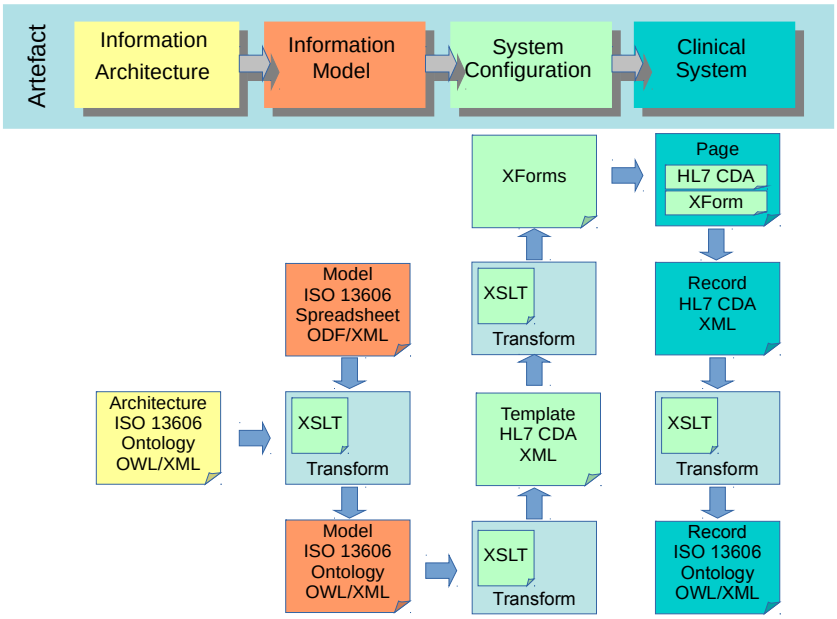

Figure 1. Model-Driven Development in cityEHR

2.1. Language Variants of the Ontology Model

The Clinical Information Models are represented as OWL/XML ontologies which capture the structure of a health record using the ISO 13606 Reference Model, whilst abstracting all terms used in the model as individual assertions in the ontology.

Support for language variants of the information models is achieved by extracting all such terms assertions in the source language, to create a 'language pack' which is translated to the target language, and then merged back into the original ontology model. The result is a model of identical structure, but with all terms expressed in the target language. Figure 2 shows a small extract of the OWL/XML representation of an information model in the source language, with the language pack used for translation to the target language.

2.2. Language Variants of Information Directories

Clinical Information Directories are held as collections of ISO 13606 Entries or Elements, represented in HL7 CDA format, so that an Entry or Element can be inserted

directly into a Composition in the patient record (represented as HL7 CDA) when selected by an EHR user.

Since the directories are designed to support localization (and so are maintained by local users), some do not require internationalization. For example, lists of clinics, primary care practices or staff roles are created locally, in the local language.

Language variants of some directories are required, however. In particular, directories of elements are used to define the 'boilerplate' text in clinical letters and whilst the structure of these letters is fixed, the text is localized and may require translation if (for example) copies are sent to clinical professionals and patients in both the official language of the EHR and the language spoken by the patient.

These language variants are supported by maintaining parallel instances of the directory, with the same structure, in multiple languages. The displayName attributes in the HL7 CDA are translated from the source language to each of the target languages.

```
<Declaration>
  <NamedIndividual IRI="#CityEHR:Term:CurrentEstimatedCalciumIntake"/>
</Declaration>
<ClassAssertion>
  <Class IRI="#CityEHR:Term"/>
  <NamedIndividual IRI="#CityEHR:Term:CurrentEstimatedCalciumIntake"/>
</ClassAssertion>
<ObjectPropertyAssertion>
  <ObjectProperty IRI="#hasDisplayName"/>
  <NamedIndividual IRI="#ISO-13606:Entry:CurrentCalciumIntake"/>
  <NamedIndividual IRI="#CityEHR:Term:CurrentEstimatedCalciumIntake"/>
</ObjectPropertyAssertion>
<DataPropertyAssertion>
  <DataProperty IRI="#hasValue"/>
  <NamedIndividual IRI="#CityEHR:Term:CurrentEstimatedCalciumIntake"/>
  <Literal xml:lang="en-gb" datatypeIRI="&amp;rdf;PlainLiteral">
   Current Estimated Calcium Intake
  </Literal>
</DataPropertyAssertion>
```

```
<DataPropertyAssertion>
  <DataProperty IRI="#hasValue"/>
  <NamedIndividual IRI="#CityEHR:Term:CurrentEstimatedCalciumIntake"/>
  <Literal xml:lang="cmn-cn" datatypeIRI="&amp;rdf;PlainLiteral">
   当前估计的钙摄入量
  </Literal>
</DataPropertyAssertion>
```

Figure 2. Part of a Language Pack for Translation of an Ontology Model from English to Mandarin

2.3. Language Variants of Configuration

Configuration files for cityEHR are maintained for both the overall system (system configuration) and for each clinical application deployed (application configuration). These configurations are held as XML files, using a custom vocabulary in which all text terms are held in displayName attributes (mirroring the markup of HL7 CDA). Language variants are maintained by extracting all displayNames from the source language configuration to create a 'language pack' into which equivalent terms can be inserted in each of the supported target languages. The language pack can then be merged with the source language configuration, replacing any displayNames with their translation in the target language.

3. Results

The cityEHR was used to create a clinical information model for a Fracture Liaison Service (FLS); this was the second generation of a model for the FLS, following an implementation which has been running at the Nuffield Orthopaedic Centre in Oxford since 2014. The model, directories and configuration files were then translated from English to Spanish and Mandarin, using tools in the cityEHR user interface designed to edit and maintain the XML structures for the language variants described previously. Based on that experience, the tools were enhanced in preparation for wider use.

At an Internationalization Workshop held at the University of Oxford in December 2021, delegates used the tools to create variants of the FLS EHR system in different target languages: German, French, Portuguese, Arabic, Farsi, Urdu and Somali.

4. Conclusions

Experience using the tools to create internationalized variants of a model-driven EHR show that the information architecture supporting language variants is sound and that the overall modeling approach is feasible, though enhancement of the tools is required.

The main issue encountered during the Internationalization Workshop was the difficulty in translating abstracted terms, where the context of the term in the EHR was not provided at the time of translation; only when the functioning EHR was generated from the model, did the context of the translation become apparent. This can be addressed by providing additional contextual information in the translation tooling.

Another common issue concerned the word order of translations, where the information model defines text strings (in clinical correspondence, for example) which are calculated from concatenation of boilerplate text in a directory and data from the patient record. To address this issue, the information architecture was extended to support strings specified as 'parsed text', rather than calculated strings. The parsed text contains both fixed text strings and references to variables, so that the entire text can be translated with the position of the variables moved as required by the target language.

References

[1] Beale, T., 2002, November. Archetypes: Constraint-based domain models for future-proof information systems. In OOPSLA 2002 workshop on behavioural semantics (Vol. 105, pp. 1-69).

[2] Kulkarni V, Reddy S. Separation of concerns in model-driven development. IEEE software. 2003 Sep 15;20(5):64-9.

[3] Kalra, D., Beale, T. and Heard, S., 2005. The openEHR foundation. Studies in health technology and informatics, 115, pp.153-173.

[4] Leslie H. OpenEHR archetype use and reuse within multilingual clinical data sets: Case study. Journal of medical Internet research. 2020 Nov 2;22(11):e23361.

[5] Min L, Tian Q, Lu X, Duan H. Modeling EHR with the openEHR approach: an exploratory study in China. BMC medical informatics and decision making. 2018 Dec;18(1):1-5.

[6] Chelsom JJ, Summers R, Pande I, Gaywood I. Ontology-driven development of a clinical research information system. In2011 24th International Symposium on Computer-Based Medical Systems (CBMS) 2011 Jun 27 (pp. 1-4). IEEE.

[7] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo A. HL7 clinical document architecture, release 2. Journal of the American Medical Informatics Association. 2006 Jan 1;13(1)

[8] World Wide Web Consortium. OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004. Available at: http://www.w3.org/T R/owl-features.