Supporting Multiple Clinical Perspectives on a Patient-Centred Record Using Ontology Models

John J Chelsom[1], Ian Gaywood[2], Ira Pande[2]

1. Centre for Health Informatics, City University, London, UK. john.chelsom.1@city.ac.uk

2. Nottingham University Hospitals NHS Trust

Abstract

Multi-disciplinary shared care is based around a single, patient-centred health record. A key driver for storing that record electronically is the need to gather data once (for clinical care) and to reuse it for secondary purposes, including clinical studies. However, physicians working in different specialties may have different perspectives on that record, both when entering new data for clinical use and when reusing those data in clinical studies. The ORCHID classification scheme is an ontology-based model which supports multiple, simultaneous clinical perspectives yet allows data to be stored as standard HL7 CDA documents in an immutable, patient-centred record. This paper describes the basic mechanisms used to support those multiple perspectives and the solution to specific problems of recording diagnosis with co-morbidities and recording different levels of detail in disease phenotypes.

Introduction

Stratified medicine requires that we move from treating the 'average' patient to making greater use of detailed characteristics of individuals and their diseases to inform clinical decisions (Hamburg and Collins 2010). This approach cannot succeed without the development of Electronic Health Records (EHR) which allow users to organise, aggregate, analyse and share the enormous volume of data generated in routine clinical care.

EHRs are, in turn, a powerful research resource both as standalone large scale data collections and in conjunction with data from tissue biobanks.

The ORCHID classification scheme is designed to enable the classification and coding of diagnoses in ways which allow physicians to develop and impose their own views on the classification of disease, whilst retaining any existing clinical codes provided by established systems such as SNOMED-CT and ICD-10.

ORCHID was developed by physicians at Nottingham University Hospitals NHS Trust and has been implemented in an EHR (of the same name) which allows data to be gathered in routine clinical encounters and then used in clinical studies, linked to data from the local biobank.

The ORCHID system uses the cityEHR open source EHR developed at City University, London, to provide both the ontology architecture of the base data dictionary in accordance with the ISO-13606 standard and the storage of clinical data as HL7 CDA documents (Dolin et al, 2006).

The objective of ORCHID is to allow data to be gathered during routine care across a range of clinical specialties, building a patient-centred record that can then be used to find cohorts of patients for clinical studies and to answer any plausible research question. The first clinical specialties covered by ORCHID are rheumatology, nephrology, respiratory and some cancer specialties.

One challenge for ORCHID is to allow physicians from each of these specialties to define clinical information models (including data sets, views and reports) which best suit their clinical practice, whilst retaining a consistent model of a patient-centred record that spans all specialties.

This paper describes how ontology models have been used to address the general problem of providing multiple clinical perspectives onto the same patient record and how specific problems have been solved related to the specification of diagnosis and co-morbidities and to the different levels of detail of disease phenotypes required in different specialties.

Ontology Model for the Clinical Record

The first step in providing multiple clinical perspectives is to decouple the specification of the different perspectives from the clinical documents stored in the clinical record. (Chelsom et al, 2011) The ontology architecture of the

ORCHID system uses the cityEHR ontology to define a Data Dictionary based on the ISO-13606 health record structure, with additional support for the entry types defined in HL7 CDA.

Within the Data Dictionary, Elements (as per the ISO-13606 model) can be designated as taking a value from the set formed by the leaf nodes in a classification hierarchy; the architecture of these hierarchies follows the ORCHID model described below. Thus the overall ontology architecture of ORCHID is a combination of the two ontologies - one representing the basic structure of the health record and its data dictionary, the other representing the hierarchical classification of data.

Using this architecture, physicians develop an information model for their clinical specialty. These models are themselves ontologies which use the axioms defined in the architecture and are represented in OWL/XML syntax (W3C Owl Working Group, 2009)

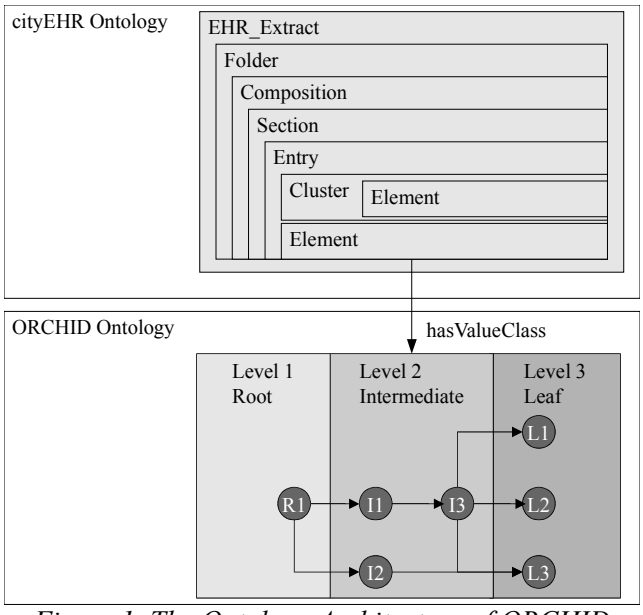

Figure 1. The Ontology Architecture of ORCHID

Hierarchical Classification of Clinical Data

The ORCHID ontology architecture provides an additional dimension to the ontology model of the cityEHR Data Dictionary by allowing a three-level classification of clinical data sets. Originally used to model hierarchies of diagnoses, the ORCHID model is now also applied to medications, laboratory test results and any other data sets can usefully be classified.

The power of ORCHID comes from its use of hierarchies and core data sets (CDS) which combine to produce highly detailed and searchable patient phenotypes. The hierarchies are generalisation hierarchies which allow multiple parents (and so are Directed Acyclic Graphs) allowing all characteristics and associations of lowest level entities to be represented.

Three levels of hierarchy are defined by the possession of parent or child nodes for any given node. Level 1 nodes describe broad concepts, typically related to a single specialty, and have no parent nodes (i.e. are root nodes). Level 3 nodes describe detailed concepts, typically an everyday diagnostic term, and have no child nodes (i.e. are leaf nodes). All other (intermediate) nodes are designated Level 2 and have at least one parent and at least one child. As there may be several levels of detail between Level 1 and Level 3 concepts, Level 2 may be multilayered.

The nodes at Level 3 form a complete and distinct set of diagnoses, and it is these nodes that are used to record the diagnosis in the patient record.

Level 3 of an ORCHID hierarchy can capture individual diagnoses but cannot efficiently capture the more detailed characteristics which define an individual's instance of that disease. This is partly because there may be many such characteristics but more because they can occur in a very large number of combinations.

The solution is to collect all of those characteristics into Core Data Sets. In the context of diagnoses, an item can be included in a disease's core data set if it says something about prognosis, severity, treatment selection or response. Examples might include lifestyle data (e.g. smoking status), clinical characteristics (e.g. weight loss), laboratory data (e.g. antibody status) or any other defining characteristics. Core Data Sets are developed for all but the simplest diseases.

In the ontology model, nodes are asserted as individuals belonging to one of the Level 1, Level 2 or Level 3 classes and the classification is made using the typeOf (inverse hasType) object property. Each Core Data Set is defined as an Entry (as per ISO-13606) in the Data Dictionary, with each item in the set defined as an Element. By this means, any clinical observation that may be gathered elsewhere in the patient record can be included in a Core Data Set.

The combination of a Directed Acyclic Graph and Core Data Sets associated with nodes in the graph is similar in concept to the model used in the Gene Ontology (Harris et al, 2004). Indeed, the interactive tools used to support the ORCHID model bear some resemblance to the tools used to support the Gene Ontology (Sealfon et al, 2006).

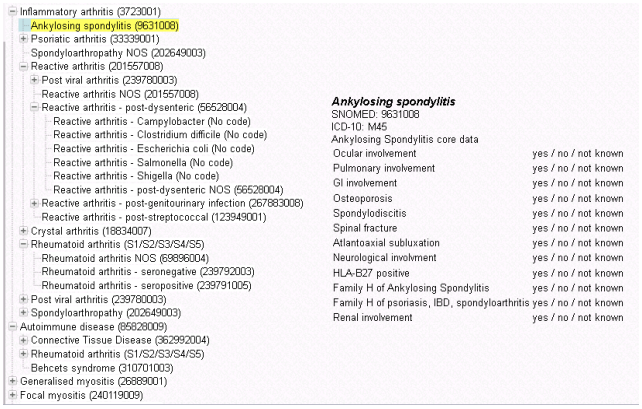

Figure 2. ORCHID Diagnosis Hierarchy with CDS

Combining Clinical Perspectives

By its very definition, an ontology is a set of assertions which provides one perspective on the conceptualisation of a real world domain (Genesereth and Nilsson, 1987). Two or more ontologies, if represented using a compatible set of axioms, can be combined to provide simultaneously multiple perspectives on the same domain.

The information models developed for each specialty contain the Data Dictionary of clinical observations, data collection forms and summary views that are required by physicians working in that specialty. In addition, each model contains the ORCHID classification of diagnoses, test results, medications, etc that are of specific interest to that specialty.

A general model is also developed to cover items in the Data Dictionary, or high level concepts in the ORCHID hierarchies that are common to all specialties. The set of models can then be merged so that each specialty has three possible perspectives on the clinical record:

1. the specialty model
2. the combined specialty and common models
3. the total of all specialties and the common model

The mechanism for combining the models is relatively straightforward, since all are represented as OWL/XML and the merging of models can be made using XSLT (Kay, 2007). However, different rules are applied for each of the two types of ontology.

For the cityEHR Data Dictionary, any common components (Entry, Element, etc) must have the same definition is each merged ontology. This means that the tools used to support the development of these models across different specialties should provide access to common components, and also that a robust governance process must be followed in order to ensure consistency and consensus across specialties.

Two main rules are applied to the merging of ORCHID class hierarchies: each node is classified using a single assertion at the lowest level in which it appears across all specialties; the set of children of any node is the union of its children across all specialties.

The Co-morbidity Problem

The concept of using different ontology models for co-morbidities has been explored by Abidi (2011). For diagnosis, physicians are generally concerned with recording the primary diagnoses from their own specialty, and noting other diagnoses as co-morbidities. In the patient record, however, there should be no distinction between diagnoses purely on the basis of which specialty recorded them. or has a special interest in them.

The ORCHID model of a diagnosis assumes that all physicians will record and use the same level of detail for any given condition. Within specialties this is most usually true (or is ensured to be true through consensus of the physicians developing the model) but there will be important differences in the information needs of different specialties for many diagnoses.

Take the following simple example: A patient has rheumatoid arthritis (RA) and chronic kidney disease (CKD). To a rheumatologist, this is a patient with RA with a complex set of disease related characteristics who also has a significant co-morbidity (CKD) which must be acknowledged and taken into account when making treatment decisions but whose underlying cause and detailed features are not immediately relevant. To a nephrologist this is a patient with CKD whose cause, prognosis, likelihood of needing dialysis or renal transplantation and metabolic consequences are of paramount importance. The fact that the patient also has RA need not be elaborated any further. This polarisation of views is starker than in reality but it serves to illustrate the fact that different specialists have different information needs relating to the same patient.

One manifestation of this problem is that a rheumatologist recording the patient diagnosis may want to see an Entry for diagnosis in which they can select Elements only from that part of ORCHID diagnosis hierarchy developed for rheumatology, with any other diagnoses recorded as co-morbidities. The nephrologist may want to see the equivalent for her specialty, with CKD shown as the diagnosis and RA as a co-morbidity. In each case, the same data for diagnosis is recorded in the patient record, but the interaction with those data will be different, depending on the specialty of the clinical user.

To support this, it must be possible to define an Entry for diagnosis in the rheumatology model which has the required behaviour for that specialty and a different Entry in the nephrology model. Yet each must result in the same entry for diagnosis in the HL7 CDA stored in the patient record.

This problem is solved by allowing Entries and Elements to be defined in the Data Dictionary for a specialty that are proxies for others in the common Data Dictionary. So the Entries for rheumatologyDiagnosis and nephrologyDiagnosis defined in their respective information models are both proxies for the general Diagnosis entry defined in the common model and used to record data in the patient record. In terms of the ontology model, this is supported through the object property isRootOf (inverse hasRoot) that defines the proxy relationship (nephrologyDiagnosis isRootOf Diagnosis).

The Lupus Nephritis Problem

We use Lupus Nephritis as a specific example of a problem in representing a diagnosis both as a node in the classification hierarchy and as an element in a core data set that defines a disease phenotype.

Rheumatologists may decide to represent SLE (lupus) as the Level 3 node Systemic Lupus Erythematosus (NOS) and capture the components of an individual's lupus, such as Lupus Nephritis, as elements in a core data set. This makes sense for rheumatologists.

Renal physicians may decide to capture Lupus Nephritis at Level 3 and capture further phenotypic details in a core data set. This makes sense for nephrologists who are interested in more detailed phenotypes of Lupus Nephritis than a rheumatologist.

The problem arises when we attempt to merge the rheumatology and renal hierarchies. We now have an item, lupus nephritis, which exists both as a Level 3 diagnosis and as an element in a core data set. Furthermore, even though the intention is that they represent the same piece of information in the patient record, they are different types of information from the perspective of the two different specialties. For the neurologists the information is of the form 'Diagnosis is Lupus Nephritis' while for the rheumatologist the information is of the form 'Diagnosis is SLE' with the core data set element 'Lupus Nephritis is true'.

The solution to this problem is to allow elements in a Core Data Set which may take values from an ORCHID class hierarchy. So in the example given above, the Level 3 node Systemic Lupus Erythematosus (NOS) used by a rheumatologist will contain an Element for Lupus Nephritis which is either empty (equivalent to the value 'false') or takes the value of the Level 3 diagnosis Lupus Nephritis.

In secondary use, when the EHR is searched to find the cohort of patients with Lupus Nephritis then the diagnosis will be found regardless of whether it was recorded by a rheumatologist as an item in the Core Data Set for SLE or as a specific diagnosis by a nephrologist.

Conclusions

The ORCHID system demonstrates that ontology-based models can be developed in different clinical specialties and then combined to provide a single model of a patient centred record into which clinical data are gathered.

The specialty models are used to provide different perspectives on the patient record for clinical users, both for data collection, viewing, reporting and secondary uses such as clinical studies.

By this means, groups of physicians can develop information models that are best suited to the requirements of their own specialty, whilst retaining the ability to interact with colleagues from other specialties using a patient-centred record which is common across all. In addition, each model also carries clinical codes from standard schemes such as SNOMED-CT and ICD-10 so that the specialty perspectives are aligned with those coding schemes to whatever extent is required.

References

Abidi, SR. 2011. Ontology-based knowledge modeling to provide decision support for comorbid diseases. In Knowledge Representation for Health-Care (pp. 27-39). Springer Berlin Heidelberg.

Bouamrane M, Rector A and Hurrell M. 2010. Experience of Using OWL Ontologies for Automated Inference of Routine Pre-operative Screening Tests. In The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010,

Chelsom JJ, Pande I, Summers R and Gaywood I. 2011. Ontology-driven development of a clinical research information system. 24th International Symposium on Computer-Based Medical Systems, Bristol. June 27-June 30 2011.

Dolin RH, Alschuler L, Boyer S, Beebe C, Gehlen FM, Biron PV, and Shabo A. 2006. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006;13(1):30-9.

Genesereth MR and Nilsson N. 1987. Logical Foundations of Artificial Intelligence. 1987 Morgan Kaufmann Publishers: San Mateo, CA.

Hamburg MA and Collins FS. 2010. The path to personalized medicine. N Engl J Med 2010;363:301-4.

Harris, MA, Clar J, Ireland A, Lomax J, Ashburner M., Foulger R, et al 2004. The Gene Ontology (GO) database and informatics resource. Nucleic acids research, 32(Database issue), D258.

Kay M. (ed) 2007. XSL Transformations (XSLT) Version 2.0. W3C Recommendation 23 January 2007.

Sealfon, R., Hibbs, M., Huttenhower, C., Myers, C., and Troyanskaya, O. 2006. GOLEM: an interactive graph-based gene-ontology navigation and analysis tool. BMC bioinformatics, 7(1), 443.

W3C Owl Working Group. 2012. OWL 2 Web Ontology Language. Document Overview (Second Edition) W3C Recommendation 11 December 2012.