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

John J Chelsom¹, Ian Gaywood², Ira Pande²

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 in use at the Nottingham University Hospitals NHS Trust in the UK, 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.

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. Its key driver is the need to take data gathered during routine clinical care and organise it in ways which can then be used to provide the detailed clinical phenotyping required for stratified medicine and for linkage to data collected for the Nottingham Health Science Biobank (NHSB).

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 (CEN/TC 251, 2005) 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 with any combination of shared attributes for clinical studies and to answer any plausible research question. The first clinical specialties covered by ORCHID are rheumatology, hepatobiliary, respiratory and some cancer specialties.

One challenge for ORCHID is to allow physicians from different 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.

Hence the cityEHR ontology model, following ISO-13606, has classes for Folders (representing the context of care), Compositions (data entry forms, summary views, letters), Sections, Entries, Clusters (of Elements) and Elements. This ontology model is then used to create the Data Dictionary for each clinical specialty, which is transformed to a set of HL7 CDA templates and other supporting XML resources for use in the EHR.

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. This provides the primary link between the two ontology models, which can be developed separately and then combined to produce the complete model for clinical use in the EHR.

Thus the overall ontology architecture the ORCHID EHR is a combination of the two ontologies - one representing the basic structure of the health record and its Data Dictionary (cityEHR), the other representing the hierarchical classification of data (ORCHID).

Using this architecture, physicians develop an information model for their clinical specialty. Each model is an ontology which uses the axioms defined in the overall architecture and are represented in OWL/XML syntax (W3C Owl Working Group, 2009).

A user (lead clinician) from each specialty iteratively develops an information model for their specialty in conjunction with the lead developers and a core user group, guided by the editorial rules governing the structure and terminology of ORCHID hierarchies. For these tasks, clinicians use open source spreadsheet and graphing tools to create the models, which are then stored in XML format and transformed to OWL/XML ontologies using XSLT. By this means, clinicians are able to develop the models using familiar tools, shielded from the complexity of the underlying ontology representation (Chelsom et al, 2012).

The issues inherent in the distributed development of an ontology model are well documented and could be addressed using central knowledge servers and repositories (Farquhar et al, 1995; Garde et al, 2007). With ORCHID, we address most issues through robust governance of the development process, with any inconsistencies between specialties identified at the time the ontologies are merged.

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 where classification is useful.

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 the 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 have no parents (i.e. are root nodes). They provide the entry points to the classification, describing broad concepts, typically related to a single specialty. Level 3 nodes describe detailed concepts, typically an everyday diagnostic term, and have no child nodes (i.e. are leaf nodes). 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.

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.

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; representing these as nodes in the hierarchy produces an unmanageable data set, in which most leaf nodes do not correspond to recognised diagnoses.

The solution is to collect the detailed characteristics of Level 3 nodes into what we term Core Data Sets. In the context of diagnoses, an item is 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. This provides a second link with the cityEHR Data Dictionary and means that any clinical observation 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 Hierarchy with Core Data Set](image)

The power of multilevel hierarchical ontologies allied to Core Data Sets lies in their ability to facilitate searches for patient cohorts at any desired level of detail, in real time, and without the need for technical insight or a knowledge of the underlying ORCHID hierarchy.

Whilst it is often sufficient to identify a cohort of patients with the same disease (rheumatoid arthritis, for example), it can be equally important to identify a broader group of patients, even crossing specialty boundaries (all patients with autoimmune disease, for example). This can be achieved by searching for diagnoses at Level 2 in the hierarchy; the actual search is then made for any descendant Level 3 nodes that have been recorded.

At a much finer level of detail, we may want to retrieve a subset of patients from among a single disease population. In this case we search for the disease at Level 3 of a diagnosis hierarchy and then further refine the selection by specifying the presence or absence of items from among the Core Data Set for that disease. An example of such a search would be patients with rheumatoid arthritis who have rheumatoid factor, antinuclear antibodies and inflammatory disease but who do not have anti-CCP antibodies or inflammatory lung disease.

The searches are made in the XML record, using the full longitudinal data set which allows complex, time-dependent searches (for example, find all patients diagnosed with rheumatoid arthritis who aged 70 or less, who have suffered a fall within the last year).

Although ORCHID has been developed in the UK National Health Service (NHS), it does not depend on any feature unique to this healthcare system. Translation to another healthcare setting would require agreement on terminology or, at least, the definition of a synonym set to cater for different terms for the same concept. A very simple example would be haemoglobin (UK) and hemoglobin (US). The need to modify the data dictionary to cope with different reference ranges (e.g. haemoglobin measured in g/dl (UK) and g/l (US)), and patient identifiers (national scale NHS numbers in the UK, provider-specific identifiers in the US) does not, in any way, undermine the architecture of the ontology model.

Indeed a single model could be used to compare patient cohorts across international boundaries, since the ORCHID/cityEHR architecture carries internationally used clinical coding (ICD-10, SNOMED CT), supports synonyms and can translate between different units of measurement. Other site-specific variations can be accommodated by developing specific information models in the same way as different models for specialties within a single healthcare organisation (described below).

### 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 multiple perspectives on the same domain simultaneously.

The ORCHID/cityEHR ontology model supports the
concept of the 'context of care' through the use of the ISO-13606 Folder, which was designed for that purpose. The clinical context of any recorded data is preserved through the HL7 CDA structure of the record. However, both the context of care and the clinical context of recorded data are patient-centred concepts that support shared care through the patient-centred record. Different members of a multi-disciplinary shared care team may also have different clinical perspectives when working within the same context of care and with the same patient record.

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 (common) 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 is then merged so that each specialty has access to three variations of the overall model:
1. their own specialty model
2. the combined specialty and common models
3. the total of all specialties and the common model

The specialty model is used in scenarios where the constraints of that specialty apply, for example in recording a primary diagnosis. The total model is used when a cross-specialty perspective is required, for example when recording a co-morbidity. The combined specialty/common model is used to support scenarios that are common to all specialties, for example the use of common assessment questionnaires or letters.

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 in each merged ontology. This means that the spreadsheet tools used to support the development of these models across different specialties must be configured to provide access to the common (shared) components, and also that a robust governance process must be followed in order to ensure consistency and consensus across specialties.

When cityEHR ontologies are merged, an error is flagged if a component is defined differently in any of the different specialty models and the merge is aborted.

Three main rules are applied to the merging of ORCHID class hierarchies.
1. Duplicate nodes with the same identity are assumed to designate the same concept.
2. Each node is classified as Level 1, 2 or 3 using a single assertion at the lowest level in which it appears across all specialties; so, for example, a node that appears at Level 2 in one hierarchy and Level 3 in another is asserted as being a Level 3 node in the combined hierarchy.
3. The set of children of any node is the union of its children across all specialties.

The issues in merging ontologies in ORCHID are less complex than the general case addressed by other tools, for example the PROMPT tool (Noy and Musen, 2002).

Merging in ORCHID takes account of the known semantics of the ontology architecture. For the cityEHR ontology, the set of axioms that define individuals in the ISO-13606 model must be identical for the merge to proceed; unlike PROMPT, violations of this constraint are flagged in the merging process but must be amended in the ontologies before the process is restarted, rather than allowing amendments during the merging process itself.

For the ORCHID class hierarchies, the merge process only fails if any Core Data Sets are defined differently (these are modeled with the ISO-13606 Entry and Element) or if the resulting merged graph is not acyclic.

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.

Consider 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, with its associated behavior. Yet each must result in the same structure for the Entry of 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 in the specialty Rheumatology model the Entry for `rheumDiagnosis` is a proxy for the diagnosis Entry that is defined in the common model and is 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 (rheumDiagnosis isRootOf diagnosis).

In the HL7 CDA stored in the patient record, the result of using `rheumDiagnosis` as the root of the common diagnosis Entry and Element is shown in Figure 3 (which excludes some XML elements and attributes not relevant to this example).

```xml
<entry>
  <observation>
    <typeid root="coidEHR"
      extension="#HL7-CDA:Observation"/>
    <id root="#ISO-13606:Entry:rheumDiagnosis"
      extension="#ISO-13606:Entry:diagnosis"/>
    <value root="#ISO13606:Element:rheumDiagnosis"
      extension="#ISO-13606:Element:diagnosis"
      value="#ORCHID:Class:Diagnosis:RANGE" orchid:coreDataSet="#ISO-13606:Entry:RACD"/>
  </observation>
</entry>
```

**Figure 3. HL7 CDA Entry for Diagnosis**

In the CDA, the XML elements 'entry' and 'value' represent the Entry and Element in the ISO-13606 model. The root attribute on the id element designates the root of the Entry; the extension attribute designates the actual identity of the Entry. Similarly, the root attribute on the value element designates the root of the Element and the extension attribute designates its actual identity.

When this CDA is manipulated in the user interface of the ORCHID EHR system, the root attributes are used to determine the behaviour of the interface, allowing the user to select a diagnosis from the correct ORCHID class hierarchy. In secondary use, when the record is used for search and analysis, the extension attribute identifies the Entry/Element in a common way, regardless of its root.

### The Lupus Nephritis Problem

We will 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 (SLE) 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 (nephrologists) may decide to capture Lupus Nephritis as a Level 3 node and capture further phenotypic details in its 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 nephrologists the information is of the form ‘Diagnosis is Lupus Nephritis’ while for the rheumatologist the information is of the form ‘Diagnosis is Systemic Lupus Erythematosus (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 (SLE) used by a rheumatologist contains an Element for Lupus Nephritis which is either empty (equivalent to the value ‘false’) or takes the value of the Level 2 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 by a nephrologist as one of the more specific Level 3 diagnoses that are descendants of Lupus Nephritis.
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.

The advantages of using an ontology architecture to create the information models (as opposed to a relational database model, for example) are primarily associated with the ease of combining and re-purposing data. All data in ORCHID are represented as XML, throughout their life cycle; in their specification, their realisation in the runtime EHR system and their preservation in the patient record. This means that both the models and the patient data can be openly exchanged and transformed.

Common available tools (spreadsheets and graphing tools) are used to create the models, which are saved as XML and transformed to OWL/XML. Models from different specialties are then combined using XML transformations (XSLT). The combined model is then transformed again to create the template HL7 CDA documents that are used to drive data capture and presentation in the EHR. The full patient record, having been created as XML (HL7 CDA) in accordance with the underlying ontology model can then be transformed back to an ontology, in accordance with that original model.

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.


