

# Models and Inference Methods for Clinical Systems: A Principled Approach

Alan L Rector, Jeremy Rogers, Adel Taweel

Department of Computer Science, University of Manchester, Manchester, England M13 9PL

## Abstract

*Previous papers have argued for the existence of three different models in many clinical information systems – for the medical record, for inference in guidelines, and for concepts and re-usable facts. This paper presents a principled approach to deciding which information belongs in each model based on the nature of the queries or inference to be performed: necessary or contingent, open or closed world, algorithmic vs heuristic. It then discusses an important class of systems – “ontologically indexed knowledge bases” – and issues of metadata within this framework.*

## Keywords:

Computerized medical records, knowledge representation, terminology, decision support systems, standards

## Introduction

Two previous papers [1, 2] proposed the thesis that three interacting models are required for clinical systems, as shown in Figure 1. Each model describes a distinct type of information source

1. *The Information Model*: what has been observed or done – the healthcare record
2. *The Inference Model*: what should be inferred or done – decision support including guidelines, protocols and warnings
3. *The Concept Model* or “Ontology”: what is necessarily or at least prototypically true – terminology and the background common knowledge base of anatomy, physiology, etc.

Principled interfaces are required between the three models. In addition, at the centre of their interface is a process of “abstraction” by which the concrete data – the actual entries in the patient record – are linked to the other two models according to context. For example, the criteria for determining whether a patient has “Asthma” will be different for a drug alert system than for entry into a controlled clinical trial of asthma patients.

However, although the existence of the three models has gained some acceptance and is confirmed by organizational structure of groups such as HL7, there has not been principled account of the differences between them nor a clear vocabulary for deciding which items should be allocated to which model. Much of the discussion focuses on the idiosyncrasies of particular technologies rather than the principles behind the reasoning.

The need for a more principled approach is made acute by a) the emergence of formal definitions and languages for Archetypes [3] as re-usable building blocks for healthcare records and the ISO/CEN efforts on Electronic Healthcare Record Architectures [4]; b) formal models of guidelines such as GLIF<sup>1</sup>, ASBRU<sup>2</sup>, ProForma<sup>3</sup>, the HL7 guideline models, and the various efforts to formalize clinical decision making models; c) the increasing number of terminology and ontology efforts including SNOMED-CT<sup>4</sup>, the National Cancer Institute Center for Bioinformatics (NCICB)’s vocabulary resources for CaCore<sup>5</sup>, and the Gene Ontology Next Generation<sup>6</sup>.

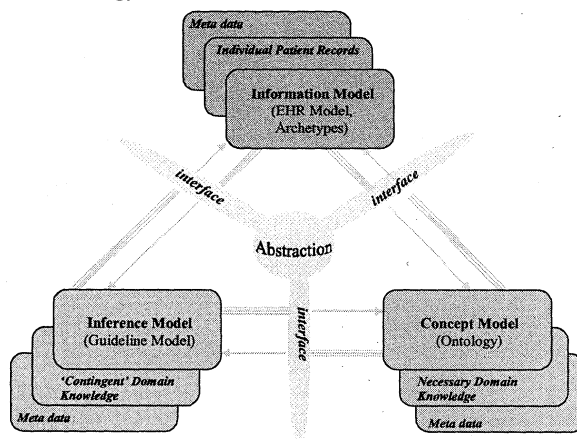


Figure 1 - Three information models for clinical systems (adapted from [1, 2])

In addition, the emergence of the Semantic Web<sup>7</sup> and the E-Science/Grid initiatives<sup>8</sup> has stressed the importance of ‘metadata’ describing the history and provenance of the models and the data they describe. Hence an extra layer of metadata has been added to each model in Figure 1 by comparison with the originals.

1. <http://www.glif.org/>
2. <http://www.asgaard.tuwien.ac.at/>
3. <http://www.openclinical.org>
4. <http://www.snomed.org>
5. <http://ncicb.nci.nih.gov/core>
6. <http://gong.man.ac.uk/>
7. <http://www.semanticweb.org>
8. <http://www.rcuk.ac.uk/escience/rcilinks.shtml>

This paper puts forward three theses, that:

- The three models are characterized by fundamental differences in the reasoning supported
- What we term an “Ontology indexed knowledge base” is a common feature in the interaction between the ontology and other two models
- There are common metadata requirements for all and that metadata is effectively contingent knowledge.

## Reasoning and models

### Vocabulary

Before discussing the different types of reasoning, it is useful to review some key properties of inference and query systems:

- *Closed vs open world* – in “closed world” reasoning, all information is assumed to be in the knowledge base or database; if information cannot be found in this specific “world”, it is assumed to be false. This behaviour is often called “negation as failure”. In “open world” reasoning, the information available is treated as a set of axioms in a logical theory to which more might always be added. Negation is taken as proof of impossibility in any “world”. This might be termed, by analogy, “negation as impossibility”<sup>1</sup>
- *Querying vs inferencing* – in querying, all reasoning is specified in the query itself. The query may be complex, but it treats the information as a set of explicit static facts – “ground clauses” in logicians’ terms; a “database” in common parlance – from which information is to be retrieved and processed but which does not contain itself any means of inference. Any inference must be either a) precomputed and inserted in the database or b) specified in the query itself.<sup>2</sup> By contrast, in inferencing, the knowledge base contains rules or axioms that allow conclusions to be inferred which are not explicitly present in the knowledge base.
- *Monotonic vs non-monotonic reasoning* – in standard logic, new information can only increase the number of conclusions that can be drawn. It can never invalidate old conclusions, *i.e.* the number of conclusions (theorems) increases monotonically<sup>3</sup>. Any closed world system is intrinsically non-monotonic with respect to negative conclusions, since new information can mean

---

1. Logicians would use the term “unsatisfiability”, *i.e.* unsatisfiable in any “world” or “model”. But beware, by “model”, logicians mean a set of facts consistent with a set of axioms. This is almost precisely the converse of what is here termed a “model”: a set of axioms (or other formalism) that constrain a set of data or facts.

2. “Deductive databases” rather blur this distinction, but can most easily be thought of in this framework as a contingent inference system tightly coupled to a database query system.

3. In standard logic, a contradiction implies anything and everything; *i.e.* in an inconsistent system all things can be proven. Hence even adding inconsistent information increases what can be concluded “monotonically”. (For an explanation see [5] or any logic textbook)

that the search no longer ‘fails’. In addition, many guideline systems allow the revision of previous conclusions on the basis of new evidence – *i.e.* they use ‘non-monotonic reasoning’.

- *Algorithmic vs Heuristic* - Algorithms are guaranteed to produce correct answers; Heuristics are rules of thumb that are useful in solving problems but cannot be guaranteed to succeed.

### Types of reasoning in health information systems

Although the above scheme might produce eight possible cells, there are only three types of reasoning relevant to this discussion:

1. *Database Querying* – simply querying of a database of facts without inferencing. No variables are allowed in the database; only in queries, and the process is algorithmic. No matter how complicated the ‘query’, at base the key information is simply retrieved from the database if present. If the information is absent, the query returns an empty set, *i.e.* the reasoning is “closed world”. For example, to query a database for a patient’s ancestors, either there needs to be an “ancestor” table in the database, or the query itself needs to specify the mechanism for retrieving an ‘ancestor’.
2. *Contingent inference* – searching for answers using sets of rules in a knowledge base about a particular individual in the specific world specified by its knowledge base and database. The rules in the knowledge base can contain variables. Reasoning can be either “backwards chaining” – *e.g.* Prolog, eMycin, etc. - or “forwards chaining” – OPS5, CLIPS, JESS, etc. To follow the above example, the rules for “ancestor” might be part of the knowledge base itself, so that the set of ancestors for an individual could be inferred without a complex query. Reasoning is “closed world” but often heuristic. Anyone about whom there was no information linking them to the individual in question would be treated as a “non-ancestor”. Many contingent inference systems also include mechanisms for “belief revision” or support for other forms of non-monotonic reasoning.
3. *Necessary inference* – finding what is *necessarily* true of any individual of a given type in any “world” consistent with the system’s axioms. Most first order logic (FoL) and all description logic reasoners as used in GALEN, SNOMED-RT/CT, and OWL-DL use necessary inference. Reasoning is “open world”, monotonic, and algorithmic. As in contingent reasoning, the axioms concerning “ancestor” might be in the knowledge base, so that the class of all of a person’s ancestors could be inferred. However, the class of “non-ancestors” would include only those who were provably not ancestors. Individuals about whom no information was available would be treated as neither ancestors nor non-ancestors.

### Reasoning in the three models

The central claim of this paper is that each type of reasoning applies to the knowledge specified by one of the three models:

1. *Queries* are used to extract information from *patient records*. Ultimately the patient record is simply a collection of facts about what carers have heard, seen, thought, and done. All that can be done is to query it. For example “On 02/03/02 Dr. Smith reported that John has diabetes”; “On date 03/04/02 Dr Jones reported that John’s diabetic control is poor”.
2. *Contingent inference* is used to make decisions based on *rules and contingent fact* about specific patients, often in clinical guidelines. Typically, the rules are highly dependent on circumstances and subject to relatively rapid change and disagreement as knowledge and practice changes. It is impractical to perform much inference in advance because relatively little can be inferred until the circumstances of an individual patient are known. Many rules are heuristic and so their results must be tested in each case. A typical rule might be: “If diabetes is poorly controlled, try increasing the insulin dose”.
3. *Necessary inference* is used to reason about *terminology and necessary domain knowledge*. The knowledge involved is either definitional – e.g. “pneumonitis” is defined as an “inflammation of the lungs” – or so deeply embedded in our common understanding of the world that we treat it as definitional – e.g. “diabetes is a kind of metabolic disorder”. The statements – or “axioms” – are true of all patients in all worlds consistent with our current understanding. Inference is algorithmic and guaranteed to succeed. Therefore, it is useful to perform much inference in advance since the inferences will apply to all patients whatever the circumstances – e.g. to infer the *is-kind-of* hierarchy amongst the concepts defined.

## Consequences and crossovers

### Abstraction is inference

An immediate consequence of the above discussion is to place the “abstraction” bubble in Figure 1 clearly in the category of “Contingent inference”. Whilst it may be appropriate, even vital, to hive it off as a separate module, it clearly requires more than simple database queries but equally does not involve universals based on definition or the fundamentals of our common conceptualisation of medicine. Hence it is likely to use heuristics, rules and contingent inference.

### “Ontology indexed knowledge bases” and re-use

The split between necessary definitional knowledge and contingent knowledge does not always fall naturally in terms of the development process. There is much that is ‘contingent’ which is nonetheless stable and re-usable across numerous applications – e.g. the uses and licensing status of drugs, the list of protocols applicable to a disease in a given hospital, the clinical significance of laboratory results, etc. Frequently the natural split in terms of labour and software architecture is between the “re-usable knowledge sources”, which are presumed to be general, and the individual guidelines, services, procedures, or messages, which are presumed to be specific. Authors of specific guidelines would like all the general re-usable information stored in the same place. Furthermore, as a matter of good software engi-

neering they would like to make the default “fail-safe” behaviour to be inherited unless over-ridden. Hence they want this information kept together with the universal information in the ontology.

In fact, ontologies make indexes to such re-usable contingent information extremely efficient. Using the ontology as a “conceptual coat rack” on which to hang other contingent knowledge often results in major simplifications[6, 7]

However, this entails contingent knowledge to be queried under a closed world hypothesis rather than necessary knowledge to be reasoned about under an open world hypothesis. The difference is not ‘academic’. Using the wrong reasoning mode gets the wrong answer! For example, a query for “drugs used but not licensed for the treatment of nausea in chemotherapy” should return all those drugs for which no license is listed, not just those for whom it has been specifically stated that they are not licensed. Otherwise it would be necessary to state all of the non-licensing explicitly – a large and pointless task.

This leads to a useful test for designers to decide whether a piece of information is ‘contingent’ or ‘necessary’: “Should the absence of information be treated as false?” “Is it practical to compile all of the negative cases explicitly?” If the answer either to the first question is “yes” or to the second question “no”, then the information should be treated as “contingent”.

We term such a combined knowledge base of re-usable contingent facts organised and indexed by an ontology an “Ontology indexed knowledge base”. A key feature of GALEN’s GRail language is support for such contingent fact<sup>1</sup>s. The need for such information is a key reason for the integration of frame systems from Protégé with the new web ontology language OWL in the CO-ODE projec<sup>2</sup>t.

### Ontology as content and ontology as index

The previous section argued for the use of “ontology indexed knowledge bases” in order to keep all re-usable information together. There is a second, perhaps even more important case for their use – when providing indexing to composite objects that cannot be listed exhaustively without producing a combinatorial explosion. This use is typical of the relationship between ontologies and data structures or guidelines. In this case the ontology plays a dual role: as *index* and as *content*.

Figure 2 illustrates this mechanism. The composite notion of “Template for Renin dependent hypertension at St Stevens Hospitals for the National hypertension survey” is formed and classified logically to place it correctly in the hierarchy as shown. The structure and position of the items for the template are ‘inherited’ as in a frame system – systolic & diastolic blood pressure from “hypertension”, “serum potassium” from “renin dependent hypertension”, etc. The meaning of those items refers back to different branches (modules) of the ontology. The fact that they are all represented in one ontology, governed by consistent logical rules, guarantees that the combined notion is correctly placed. Because the ontology is “normalised” conflicts in

1. termed “extrinsics” in GRail for historical reasons  
2. <http://www.co-ode.org>

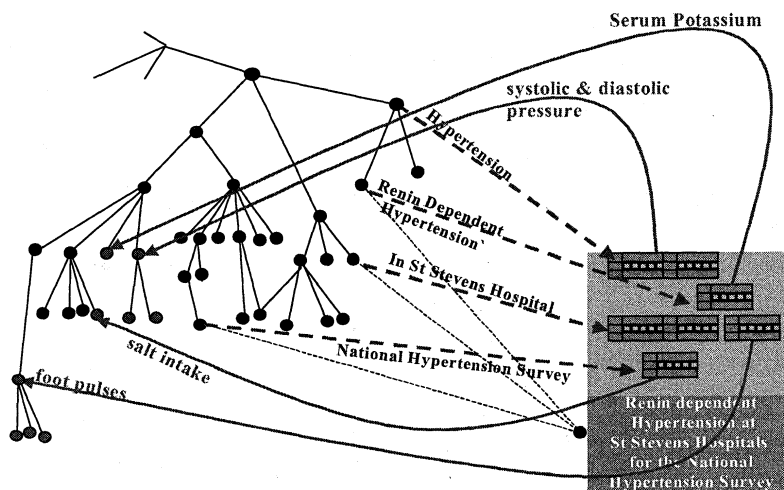


Figure 2 - Dual Role of Ontologies as Content and Index

inheritance – “Nixon diamonds” - are rare. This was the fundamental mechanism behind the PEN&PAD system [8].

## Metadata and Provenance

Metadata is traditionally described as “data about data”. We can divide metadata into two classes:

- Metadata about the representation – *e.g.* editorial information about how this information came to be in this form in the knowledge base or EHR.
- Metadata about the actual information itself – who first described a disease or disease class, whether the concept is current or outmoded, etc.

The second case, metadata about the actual information, is beyond the scope of this paper. The first case, metadata about the representation is a major feature of both Electronic Healthcare Records and most guideline systems. For example the GEM measures of guideline quality are largely concerned with the metadata – who compiled the guideline, from what authority, how thoroughly it has been tested, etc. Much of the medical record consists of metadata about who has responsibility for decisions and data.

Such metadata is almost always contingent. In fact it can be considered as simple ground statements about the data objects themselves *qua* data objects. As such, it is information which we expect to query under a closed world hypothesis rather than use in inferring classification under an open world hypothesis. For example, when using the Dublin Core<sup>1</sup>, which widely used for holding reference information for digital libraries, when asking about plays written by Gilbert but not Sullivan, I would expect to get back those on which only Gilbert’s name appears. I would not expect to have to have listed explicitly that they were not written by Sullivan. The same will be true when looking, in a parallel example, for concepts which appear in SNOMED-CT

but not ICD-10, or for entries in the record which were not made by a junior doctor. The issue is important because the role of such metadata statements controversial even within the standards debates around the new web ontology language, OWL.

## Discussion

It is frequently unclear which of the three models should contain which information. As the formalisms used in each model become more expressive, this problem becomes more common rather than less, because at least superficially, more things can be represented in any of the three models. The expressiveness of the new web ontology language, OWL, is therefore a potential hazard as well as a benefit.

Some choices will, inevitably, be pragmatic. However, in others, these criteria at least give a basis for argument. For example, consider whether the classification of patients as having rheumatoid arthritis on the basis of having at least five out of the seven criteria on the American Rheumatology Association scale. The new web ontology language, OWL, is powerful enough to express this notion<sup>2</sup>. However, is what we really want to say: “Necessarily, anyone having five of these symptoms has rheumatoid arthritis” with the same strength that we say: “Necessarily, anyone who has rheumatoid arthritis has an autoimmune disease”? Almost certainly not. International differences in the criteria for the diagnosis of rheumatoid arthritis exist precisely because we do not have a sufficient understanding to make such a strong statement. Nor do they hold without exception. It is not *contradictory* to find a patient who fails these criteria but whom we still believe to have rheumatoid arthritis, merely unusual. Hence diagnosis via such criteria is contingent reasoning and belongs in the inference model.

By contrast, consider “Idiopathic hypertension”. The notion of idiopathic hypertension really is defined by having no demon-

1. <http://dublincore.org/>

2. Awkwardly in version 1, but easily in likely subsequent versions which will include “Qualified number restrictions”

strable cause. Patients should not be considered to have “idiopathic hypertension” just because no cause is recorded, but only when they have been investigated and shown to have no plausible cause. Hence, the notion is definitional and the reasoning “open world”. Therefore, the definition of idiopathic hypertension is necessary inference and belongs in the concept model.

Finally, consider the issue of how best to handle a set of data elements – archetypes, templates, or data entry forms, etc. If there is a well defined modest number of these elements, then individual enumeration is sufficient and the use of an “Ontology indexed knowledge base” is unnecessary. If the total potential number is *combinatorially* large but the number of the actually defined is modest, then an “Ontology indexed knowledge base” is a plausible solution.

Whatever the choice, the questions: “Do I need simple queries or complex inference?”, “Is it contingent or necessary?”, “Is the reasoning closed or open world?”, “Is the indexing liable to combinatorial explosion?” provide a framework for rational discussion.

### Acknowledgements

This work supported in part under the UK E-Science Initiative by grant G0100852 from the Medical Research Council- the Clinical E-Science Framework (CLEF), the CO-ODE project funded by Joint Committee for the Support of Research and grant GR/S44686/01 from the Engineering and Physical Sciences Research Council. Special thanks to partners in the GALEN and CLEF programmes and colleagues at Stanford and HL7.

### References

- [1] Rector AL. The Interface between Information, Terminology, and Inference Models. In: Patel V, editor. Tenth World Conference on Medical and Health Informatics: Medinfo-2001; 2001; London, England; 2001. p. 246-250.
- [2] Rector AL, Johnson PD, Tu S, Wroe C, Rogers J. Interface of inference models with concept and medical record models. In: Quaglini S, Barahona P, Andreassen S, editors. Artificial Intelligence in Medicine Europe (AIME); 2001; Cascais, Portugal: Springer Verlag; 2001. p. 314-323.
- [3] Beale T. Archetypes: Constraint-based domain models for future-proof information systems. In: OOPSLA-2002 Workshop on behavioural semantics; 2002; available from [http://www.oceaninformatics.biz/publications/archetypes\\_new.pdf](http://www.oceaninformatics.biz/publications/archetypes_new.pdf); 2002.
- [4] Kalra D, Austin A, O'Connor A, Patterson D, Lloyd D, Ingram D. Design and Implementation of a Federated Health Record Server. In: Toward an Electronic Health Record Europe 2001; 2001: Medical Records Institute for the Centre for Advancement of Electronic Records Ltd.; 2001. p. 1-13.
- [5] Genesereth M, Nilsson NJ. Logical Foundations of Artificial Intelligence. Los Altos, CA: Morgan Kaufmann; 1987.
- [6] Solomon W, Wroe C, Rogers JE, Rector A. A reference terminology for drugs. Journal of the American Medical Informatics Association 1999(Fall Symposium Special Issue):152-155.

[7] Rector AL, Zanstra PE, Solomon WD, Rogers JE, Baud R, Ceusters W, et al. Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine. IEEE Transactions on Information Technology in BioMedicine 1999;2(4):229-242.

[8] Nowlan WA. Clinical workstation: Identifying clinical requirements and understanding clinical information. International Journal of Bio-Medical Computing 1994;34:85-94.

### Address for correspondence

Alan Rector, Dept of Computer Science, University of Manchester, Manchester M13 9PL, England