

Clinical Terminology: Why is it so hard?

Alan L. Rector

Medical Informatics Group, Department of Computer Science
University of Manchester, Manchester M13 9PL

Tel: +44-161-275-6188/7183 Fax: +44-161-275-6204

email: rector@cs.man.ac.uk

<http://www.cs.man.ac.uk/mig/>

1999 *Methods of Information in Medicine* 38(4):239-252

Abstract

Despite years of work, no re-usable clinical terminology has yet been demonstrated in widespread use. This paper puts forward ten reasons why developing such terminologies is hard. All stem from underestimating the change entailed in using terminology in software for 'patient centred' systems rather than for its traditional functions of statistical and financial reporting. Firstly, the increase in scale and complexity are enormous. Secondly, the resulting scale exceeds what can be managed manually with the rigour required by software, but building appropriate rigorous representations on the necessary scale is, in itself, a hard problem. Thirdly, 'clinical pragmatics' – practical data entry, presentation and retrieval for clinical tasks – must be taken into account, so that the intrinsic differences between the needs of users and the needs of software are addressed. This implies that validation of clinical terminologies must include validation in use as implemented in software.

1. Introduction

1.1 The problem of medical terminology

For at least the last decade, problems of standardising medical language and terminology have been a major concern of Medical Informatics. Sittig placed achieving a common ‘controlled vocabulary’ at the top of his list of ‘grand challenges’ for medical informatics [1]. Major efforts have been mounted by the United States National Library of Medicine in its UMLS project, the UK National Health Service and its Centre for Coding and Classification, by SNOMED International, and by the GALEN programme of the European Community.

Meanwhile, terminology standards have been a major effort for CEN Technical Committee 251 (CEN/TC251[2] and ISO TC215 [3, 4] There are major efforts on vocabulary within other standards efforts including [5], and the DICOM [6], and CorbaMed made one of its first Request For Proposals and standards the specification for a ‘Terminology Server’ [7]. In the US, the Kennedy-Kassebaum [8] amendment will mandate a national vocabulary early in a few year’s time, and the Group of Eight industrial nations efforts on health have also given prominence to problems of terminology [9].

In decision support, investigators such as Musen [10, 11] and Fox [12] have had considerable success in developing re-usable inference and problem solving methodologies (once thought to be the more difficult problem). However, re-using the ‘ontologies’ or terminologies needed for the re-usable problem solving methods has been much less successful and is now a major constraint on further progress.

A large fraction of recent major meetings on medical informatics was devoted to problems in the area of language and terminology. Most recently the Convergent Terms Project [13] and its successor SNOMED-RT have embarked on a multi-million dollar effort to produce a formal medical representation [14].

In the face of such concentrated effort, such a seemingly straightforward problem ought to have been solved. Judging by the scale of the continuing efforts, it has not been, and there remains scepticism as to whether the current efforts will succeed in providing the needed solutions. The difficulty is made more puzzling because special purpose solutions to small scale problems are easy – dozens of small special purpose vocabularies exist, most of which were developed relatively quickly.

Given that major decisions are pending and major efforts coming to maturity, it is an appropriate time to step back and look at the reasons that medical terminologies are hard. More importantly, it is important to be sure we can answer the question: ‘How will we know if we have succeeded?’.

This paper investigates the requirements and conflicts which make developing such a comprehensive re-usable terminology for patient-centred systems hard. In setting out these issues, this paper aims to broaden and deepen the debate on the structure of terminologies and, in particular, to understand and elaborate the desiderata set out by Cimino in his influential paper [15].

1.2 Assumptions

This paper makes three assumptions about the context in which terminologies are being used:

1. *The purpose of clinical terminology is to support clinical software.* Were it not for the desire to build Electronic Healthcare Records and computer assisted decision support, quality assurance, and information management systems, developing elaborate clinical terminologies would be at most a curious philosophical pastime rather than a vital engineering priority.
2. *All terminologies will have to support conversion to existing reporting and epidemiological coding schemes* such as ICD 9/10/9-CM, CPT4, OPCS4, etc. and that these systems are, in themselves, complex structures with complex organisation and integrity rules which must be respected.
3. *All terminologies will need to be multilingual* because clinical professionals need to see their information in their own language.

The goal therefore is to develop multilingual terminologies suitable for use in software and compatible with existing practice and reporting systems which meet the aspirations below.

1.3 Definitions: What is ‘terminological’

What is the scope of ‘terminology’ and ‘terminological reasoning?’ What is the role of the terminology in clinical information systems? For this paper we shall propose a relatively simple definition:

Clinical terminology concerns the meaning, expression, and use of concepts in statements in the medical record or other clinical information system.

Terminological reasoning is that reasoning which can be performed on the basis of the classification, relations and comparison of isolated concepts from a medical record or information system

An important part of the purpose of this definition is to identify what is *not* terminological inference. From this definition, the inference from ‘*This patient has diabetes*’ to ‘*This patient has an endocrine disease*’ is terminological because it involves just the meaning of the concept ‘*Diabetes*’ and its classification. The inference from ‘*This patient has repeated abnormal glucose tolerance tests*’ to ‘*This patient as diabetes*’ is a diagnostic inference beyond the scope of simple terminology since it involves several statements and more inference than the simple classification of the concepts ‘*abnormal glucose tolerance test*’ and ‘*diabetes*’. Similarly, the inference from ‘*This patient has Rheumatoid Arthritis*’ to ‘*This patient has an auto-immune disorder*’ we would take as terminological; the inference from ‘*This patient shows five out of seven criteria on the American Rheumatology Association scale*’ to ‘*This patient has Rheumatoid Arthritis*’, we would take as general diagnostic inference beyond the scope of terminology.

The above leaves the notion of ‘concept’ undefined. Perhaps the simplest definition comes from ISO TC37: “a concept is a unit of thought” [16]. (In most situations, we can think of concepts as ideas represented by noun phrases, although the flexibility of language means that almost anything which can be expressed as a noun phrase can also be expressed using other constructs.)

1.4 Aspirations

1.4.1 Goals: ‘Patient Based Systems’ and the re-use of Information

Most current development in medical informatics aims at systems in which information is shared for a variety of purposes. There are at least four underlying assumptions which can be found in work as diverse as the NHS IM&T [17] and US Government pronouncements and the pronouncements of organisations such as the Computer Based Patient Records Institute [18].

1. *All patients should have Electronic Patient Records.*
2. *All information should be derived from information entered at the point of care into Electronic Patient Records; hence information must be capable of being used both for clinical care and secondary purposes. Furthermore, the goal is usually that information collection be seamless – i.e. that healthcare professionals should enter information in the patient record without needing to be aware that the information will also be used for secondary purposes.*
3. *Electronically shared information should be a key part of co-ordinating shared patient care amongst different institutions, professions, and specialities; hence the same information should be re-usable for different types of care by different professionals*
4. *Care will increasingly be organised according to agreed guidelines, and that such guidelines can only be made practical when delivered electronically integrated with electronic patient records.*
5. *Information will increasingly be delivered electronically via the Web and related technologies, filtered and relevant to particular patient situations, implying that indexing and access to knowledge resources will need to be co-ordinated with patient care systems.*

1.4.2 Information types, Tasks, and Users: “What’s it for?”

Curiously, despite the amount of literature on medical terminologies, there is relatively little which specifies the specific tasks which the terminology must perform or facilitate, and only slightly more which discusses the overall activities and users to be served. The following list is culled from a variety of sources and the implications of various papers. We divide the information into the types of information to be collected, the tasks to be performed with that information, and the potential users of that information.

Information types::

1. *Information on individual patients* – medical records
2. *Information on populations of patients* – aggregated data for epidemiology, public health, and quality assurance
3. *Information on institutions and the health care system* – information for planning, management, and remuneration

4. *Information on the current state of knowledge of best medical practice* – knowledge management and decision support in its widest sense.

Primary Tasks:

1. *Entering data about patients* – fast, easy and intuitive enough to be used routinely during or immediately patient consultations
2. *Presenting information about individual patients* – getting information from the medical record
3. *Querying and retrieving information* about populations of patients
4. *Sharing and integrating information* from different applications, medical records, and decision support systems

Secondary tasks

1. *Navigating and browsing* through information – either locally or on the web
2. *Authoring knowledge* – either static knowledge for browsing or dynamic decision support
3. *Indexing knowledge* – both general medical knowledge and information about individual patients
4. *Analysing and generating natural language* – for international use in multiple languages according to local usage and preferences

Users:

1. *Doctors, nurses, physiotherapists, clerks, secretaries* and others engaged in performing and recording direct patient care
2. *Managers, planners, and financial staff* engaged in managing health care enterprises
3. *Researchers*, public health staff, and those engaged in quality assurance
4. *Patients and the general public*, who are increasingly demanding access to their medical records and general medical information
5. *Health information management and communications professionals*
6. *Health care enterprises* – although strictly speaking not ‘users’, entire healthcare enterprises have needs distinct from the needs of any individual working in them. The enterprise wide needs for a coherent approach to information supporting information sharing, integration, security, and management makes it easiest to treat the enterprise as a whole as a special kind of user for purposes of analysis.

Such a protean list of aspirations must give us pause? Is it realistically possible? We must face the fact that *there is as yet no proof that a general re-usable terminology serving all of the aspirations for clinical information systems is possible.*

2. The difficulties: ‘Why it is so hard’

This paper puts forward ten reasons why clinical terminology for software is hard.

1. The scale and the multiplicity activities tasks and users it is expected to serve is vast.
2. Conflicts between the needs of users and the requirements for rigorously developed software must be reconciled
3. The complexity of clinical pragmatics – support for practical use for data entry, browsing, and retrieval – and the need for testing the pragmatics of terminologies implemented in software.
4. Separating language and concept representation is difficult and has often been inadequate.
5. Pragmatic clinical conventions often do not conform to general logical or linguistic paradigms.
6. Both defining formalisms for clinical concept representation and populating them with clinical knowledge or ‘ontologies’ are hard – and that their difficulty has often been underestimated.
7. Determining and achieving the appropriate level of clinical consensus is hard and requires that the terminology be open ended and allow local tailoring.
8. The structure idiosyncrasies of existing conventional coding and classification systems must be addressed
9. The terminology must be co-ordinated and coherent with medical record and messaging models and standards

10. Change must be managed, and it must be managed without corrupting information already recorded in medical records.

The remainder of this paper discusses these issues and their ramifications.

2.1 Scale and diversity of use

The first reason clinical terminology is hard is therefore its vast scale and the multiplicity of potential activities, tasks and users it is expected to serve.

The lists in Section 1.3 should give pause for thought. The total combination from the products of information types, tasks and users is very large indeed. It suggests that meeting the needs of patient based systems may involve a radical scaling up of existing terminologies. The scope of medical knowledge across all specialities, for detailed clinical care, in particular, is orders of magnitude larger than the terminology needed to report mortality and morbidity or simple diagnostic registers in a single speciality or even general practice. Experience in scaling up systems such as ICD or The Read Codes version 1 from use for epidemiological reporting and special purpose use respectively to attempts at more general re-usable systems confirm this observation with increases from no more than 20,000 items to in excess of 250,000.

This reflects a fundamental property of digital systems: scale tends to increase exponentially. For example, if we take two kinds of burns, thermal and chemical, and 200 body locations, we get 400 codes for burns, if we add three thickness plus unspecified, we have 1600; if we add three extents plus unspecified we have 6400; add three levels of recency plus whether or not the burn is complicated by infection, and the total rises to 76800; etc. just for burns. This ‘exponential explosion’ is fundamental to any attempt to enumerate all combinations of characteristics

Scaling up by an order of magnitude or more is notoriously difficult, and often requires change in methodology and technique.

2.2 Terminologies for People and Terminologies for Machines:

The second reason that terminology is hard is that there are fundamental conflicts between the needs of users and the requirements for rigorously developed software, and this difficulty is compounded by a legacy of techniques designed to meet constraints which no longer apply.

At the heart of the difficulties is the fact that humans and machines process information very differently. To borrow the words Donald Norman, of one of the gurus of human centred design:

We are analog beings trapped in a digital world... We are compliant, flexible, tolerant. Yet we have constructed a world of machines that requires us to be rigid, fixed, intolerant ([19] chap 7; pg 135).

Perhaps the most important difference is that small changes in analogue systems usually give rise to small changes in outcome whereas small changes in digital systems often give rise to unpredictably large changes in outcome. Furthermore, analogue systems tend to scale predictably – the behaviour of a large bridge can usually be well predicted from the behaviour of a scale model . Whereas digital systems tend to scale unpredictably – a system ten times as large may require one hundred, or even one hundred thousand, times the computational resources as the smaller system. The formulae for scaling analogue systems are, with a few exceptions, simple well known laws of physics, *e.g.* the ‘square-cube law’; the laws for scaling in digital systems include some of the more counter-intuitive results of twentieth century mathematics.

To make the argument more precise, a clinical terminology must at the same time:

1. Be understandable by human healthcare professionals in their own language
2. Be usable and intuitive and fit into the healthcare professionals’ daily routine
3. Behave in a rigorously predictable way for software engineers; in particular the classification of concepts should give rise to accurate retrieval of information.

The first two requirements need compliant, flexible tolerant ‘analogue’ systems; the third requirement needs a rigid, fixed intolerant ‘digital’ system. The consequence of the assumption that ‘the purpose of clinical terminology to support clinical software’ is that the behaviour of the terminology which is relevant to that software must be rigorously defined. Yet somehow we must accommodate the flexibility required by human users.

To make matters worse, insofar as existing terminologies – ICD, SNOMED, CPT, etc.– were designed for use with computers, they were designed for machines from the era where processing power and storage were expensive, and human users were necessarily required to adapt to the limitations of software if they were to use it at all. For example, the one of the main purposes of the Read Codes (version 1) was make to make it possible to store information compactly as four-digit alphanumeric codes to fit on floppy disks with a capacity of a few hundred kilobytes. Progress in hardware means that such constraints which were once paramount are now irrelevant. We aspire to make ‘computerisation’ nearly universal for a vast range of tasks and users who have no reason to adapt to the needs of the computer, but existing approaches to terminology carry the legacy of conditions long past when there was no possibility of adapting computers to users at feasible cost.

2.3 The importance of Clinical Pragmatics

The third reason that medical terminology is hard is the complexity of clinical pragmatics – support for practical use for data entry, browsing, and retrieval – and the need for testing the pragmatics of terminologies implemented in software.

The above section separated out two user-oriented ‘analogue’ requirements. Corresponding to each is a discipline and task:

1. Clinical computational linguistics – getting the language right
2. Clinical pragmatics – getting the clinical dialogue right so that it fits into daily tasks and practice.

Clinical computational linguistics is a relatively well developed field with significant successes such as the Linguistic String Project [20], the work of Scherrer, Baud, Rassinoux and others in Geneva [21-24], and others [25, 26].

Clinical pragmatics is much less well developed, even ignored. This paper borrows the word ‘pragmatics’, the study of dialogue, but stretches it to all those aspects of the ‘user interaction’ with the information system which make the terminology ‘useful and usable’ – the structures needed to support data entry, browsing, structured retrieval, etc. The specific difficulties in clinical pragmatics are discussed in Section 3.2. However, many developers of terminologies explicitly dissociate themselves from clinical pragmatics – how the terminology in clinical systems – consigning it to ‘implementation’.

There are four problems with this position:

1. The structure of the terminology significantly affects how it can be used in dialogue. For example, whether a notion must be expressed as a single concept which must be selected or whether several concepts can be combined profoundly affects what sort of interfaces can be built. Similarly, the richness of the relations amongst concepts represented in the terminology profoundly affects how browsing and retrieval can be structured.
2. The scaling problem is at least as great for clinical pragmatics as for clinical linguistics and concept representation. There are potentially combinatorially many clinical situations requiring clinical dialogues. In a single area, it is always possible to build an effective user system and then ‘bolt on’ a terminology at the end. Across all of medicine, such a solution runs into the combinatorial explosion. A general solution for clinical pragmatics must scale. Unless the pragmatics scale, the only economically feasible activity is a ‘one-size-fits-all’ approach which too frequently suits no one. The structure of the concept system and linguistic components is critical to whether or not building a scalable dialogue system is possible.
3. The terminology cannot be evaluated without clinical pragmatics. If the goal of clinical terminology is to support clinical software, then any evaluation must include evaluation of the terminology’s performance in software. It is certainly true that any terminology can be so badly implemented as to be unusable. However, in the absence of an ‘existence proof’ of a good scalable interface using a given terminology in practical software for clinical use, there is no evidence that such systems can be built. Not to include evaluation in software as part of evaluation of terminology is analogous to not testing drugs *in vivo* but only *in vitro*. Put another way, it is to make ‘success’ rather than ‘failure’ the null hypothesis of the experimental evaluation. Nowhere else in medical research would this be considered acceptable. There are organisational and commercial problems to be overcome, but these cannot be taken as justification for assuming without evidence, sometimes in the presence of consistent evidence to the contrary, that it ‘must’ be possible to implement a given terminology so that it is usable.

4. Empirically, many vendors have chosen not to use existing terminologies and opted instead either systems developed in house or of commercial interfaces such as those of MEDCIN™ or Purkinje's as have some academic developers of systems emphasising user-oriented design [27-29]. The commercial efforts have been largely ignored by the academic community but their user communities provide *prima facie* evidence of practical clinical acceptance of the resulting systems, whatever their other defects and limitations. Similarly, developers of messaging standards, HL7, LOINC and DICOM, have tended to develop their own, special purpose, terminologies in varying degrees of co-operation with the main providers. Until proven otherwise, the fact that so many commercial vendors have chosen not to use standard terminologies must be taken as strong presumptive evidence that those terminologies are seriously flawed from the point of view of practical use in scalable systems.

3. Three fundamental disciplines: Conflicts and confusions.

To emphasise the importance of the distinctions above, we propose the following:

Hypothesis of separability: For a clinical terminology, the representation of concepts and the relations between them can and should be separated from the linguistic knowledge about how these concepts are expressed in language and the pragmatic knowledge concerning how these concepts are used in dialogues with clinical users.

Perhaps the clearest way to motivate the hypothesis is to look at their alternative criteria for correctness in each of the three disciplines. It is implausible that a single integrated system could meet them all simultaneously, and this intuition is born out by practical experience.

1. *Clinical linguistics* – expressing concepts in words which sound natural to a native (clinical) speaker of the language in question.
2. *Clinical pragmatics* – organising information in ways which correspond to how healthcare professionals expect them and in ways that facilitate their daily work
3. *Logical concept representation* – formal representation of concepts in ways which give rise to correct identification, classification, and retrieval of information in formal (computer) systems.

The criteria and needs of the three disciplines often conflict. Each requires different information and reasoning. Linguistic and concept structures may differ – “Left nephrectomy” sounds natural, but users know that it means something like “removal of the left kidney” rather than “left removal of kidney”. Names may be misleading, *e.g.* ‘Cor pulmonale’ is a condition of the lungs rather than the heart. Common usage may differ from the expectations of information retrieval, *e.g.* it may seem odd to say that ‘The fingernail is a part of the arm’, but it would be an error not to retrieve ‘evulsion of nail’ under ‘trauma to upper extremity’. Clinical pragmatics may require distinctions which have no direct correspondence to the usage of either linguistics or concept representation, *e.g.* “signs of heart disease” in a history normally includes not the entire list of all things which might be signs of heart disease, but a convenient subset of them defined by convention and practice. Similarly, ‘auscultation of the heart’ may mean include very different lists of findings at different levels of granularity in general practice by comparison with cardiology. Indeed, as Kay and Purves have pointed out, the medical record has a complex narrative structure which we capture at best partially [30]. Terminology must work within this framework.

In addition, because each of the three disciplines has developed separately, each uses different tools and techniques which are based on fundamentally different underlying principles. Integrating such systems directly is difficult, perhaps impossible, whereas designing a system around a set of communicating components with defined interfaces makes most such differences irrelevant.

The following two sections explore the consequences of separating linguistics and pragmatics from the formal representation and each other.

3.1 Clinical Linguistics and Formal Concept Representation.

The fourth reason medical terminology is hard is insufficient separation of language and concept representations.

One of the major difficulties in medical terminology has been the confusion of concepts and the words used to express those concepts, or more properly between the linguistic representations and the concept representations. The difficulties are made worse because the formalisms and language used in linguistics and concept representation are often similar.

Key effects therefore of the separation of linguistics and concept representation include:

1. *No information used to process concept representations can be derived from language in the rubrics or names of concepts.* Likewise, the linguistic information needed to express or understand the concepts in natural language – lexicons, grammars, etc – should not affect the internal workings of the concept representation system.
2. *The concept representation system should behave exactly the same way if all internal identifiers were replaced consistently by any set of meaningless symbols, i.e. the internal system is ‘digital’ in Norman’s sense (see 2.2) or, to use Simon’s well known notion [31], it is a ‘Physical Symbol System’.*
3. *‘Ambiguity’ is a phenomenon of the interaction between the linguistic and concept components in which one linguistic expression can be interpreted as more than one internal concept.*
4. *Multilingual systems can be built by attaching alternative language components to the same concept representation and pragmatic components.*

Note that the hypothesis of separability should not be confused with the assertion that everything in the text of the clinical notes can be represented formally. The clinical notes expressed in natural language will, for the foreseeable future, be richer in content and context than any formal representation of them. Similarly, the hypothesis of separability does not mean that semantic constraints from the concept representation component cannot be used by the linguistic component, only that the concept representation component cannot depend on constraints from the linguistic component. (Although linguistic tools might be used to help *construct* the concept representation initially.)

Finally, it should be noted that the hypothesis of separability is not believed by most linguists and philosophers to be true in general. It appears reasonable, even essential, for special purposes in medicine because a) there is a broad common consensus on a core model of anatomy and pathophysiology, at least amongst western doctors, b) the goals for clinical terminology though broad, fall far short of general natural language – they encompass at most straightforward statements and do not aspire to poetry, literature, or even innuendo.

3.2 Clinical Pragmatics, Clinical Linguistics and Formal Concept representation

The fifth reason that that medical terminology is hard is that clinical conventions often do not conform to the usual logical or linguistic paradigms.

Clinical usage has developed over generations. It frequently uses unusual linguistic constructs special to itself, and it often defies literal logical interpretation. Four key issues are:

1. *Clinical convention* – phrases do not literally mean what they say. Getting a definition right can be more difficult than it looks. For instance ‘Heart Valve’ would seem to mean “valve in the heart”, but under that definition the *foramen ovale* counts, since it unquestionably functions as a valve and is a structure in the heart. So also does a ‘*Star Edwards Prosthetic Valve*’, at least when installed. Most doctors expect to see a list of the four main heart valves listed under ‘*heart valve*’, so additional information on common usage beyond either linguistics or the formal concept representation is required.
2. *Clinical Expectation* – how to be comprehensive, relevant, and normative simultaneously. Most of the time, a user expects to find ‘*upper*’ and ‘*lower*’ ‘*lobes of the left lung*’ but ‘*upper*’, ‘*middle*’, and ‘*lower*’ ‘*lobes of the right lung*’. Any interface which offered ‘*left middle lobe*’ routinely would be annoying. However, there are patients with complete *situs inversus* who do indeed have a left middle lobe. Suppressing the unusual while at the same time catering for it when appropriate requires additional information.
3. *Operational meaning* – knowledge sources such as the British National Formulary often use phrases which would not appear in any natural way in a medical record, e.g. “tendency to ST prolongation” [32]. On the one hand, this concept is easy to represent; on the other it is so unlikely to be recorded as such in the medical record that its use as a ‘*contraindication*’ is operationally useless. It is a significant clinical task to find out what situations the term is intended to cover which might actually be recorded in an operational record.
5. *Special use of language in the rubrics of existing coding systems*, which often omit information from higher levels, confound ‘and’ and ‘or’, and use a variety of peculiar constructs.

3.3 Formal Concept representation: Requirements and constraints

The sixth reason that clinical terminology is hard is that both formalisms and ontologies for clinical concept representation are hard – and that their difficulty has often been underestimated.

Even once separated from the complexities of linguistics and pragmatics, formal concept representation itself is a difficult topic in itself.

Furthermore, the problem of concept representation must itself be split into two parts:

1. The representation language or formalism – the well studied candidates are the knowledge representation languages, also known as description logics, which have developed out of semantic networks via KL-ONE [33, 34] and Conceptual Graphs [35].
2. The schema and content or ‘ontology’ – the representation of medical concepts using the formalism, where there has been much less formal study though numerous significant projects outside of medicine, *e.g.* Cyc [36], the Darpa Knowledge Sharing Effort [37], and the Penman Upper Model [38].

Both are difficult.

A detailed examination of formal knowledge representation is beyond the scope of this paper, but key requirements for medical systems are discussed below.

3.3.1 Making information explicit

What is required is a rigorous logical language, or representation, capable of explicitly defining and describing concepts with criteria for when two concepts are the same, when one is a kind of the other, or how they differ.

If the strategy for patient-centred information systems is to work, information must be transformed logically from the form in which it is entered for use in patient care to the forms in which it will be re-used for other purposes. Unfortunately, standard terminologies themselves lack much of the information need for such transformations. They were designed to be used and interpreted by people using their background clinical knowledge. What limited information is present is, is expressed in the natural language in the rubrics which cannot be used by the computer system. The task, therefore, is to make the information which is implicit or hidden in the rubrics explicit in a form which can be used for computer analysis

There are four primary kinds of information are usually to be made explicit:

1. *Definitions of concepts* – sets of criteria collectively sufficient to recognise a given concept *e.g.* that ‘hepatitis’ is an inflammation of the liver. If a second concept’s definition plus description satisfies a concept’s definition, it should be classified as a kind of it.
2. *Description of concepts* – facts which must be true of a concept and which can be used to classify it but which may, or may not, be sufficient to recognise it *e.g.* that the aortic valve is part of the heart. If one concept’s definition plus description satisfies a second concept’s definition, then the first concept should be classified under the second.
3. *The criteria for classifying concepts into hierarchies.* For example, long usage and clinical knowledge make it obvious to human users why congenital conditions, respiratory conditions, and tumours should each get a separate chapter in most classifications, even though the criteria for inclusion in each is different: aetiology, anatomical location, and morphological structure respectively. However, the reason for classification is neither consistent nor explicitly represented in a form usable by computer systems. (Indeed, librarians specifically recognise the heterogeneous nature of their hierarchies by insisting that the hierarchical relation be labelled simply ‘broader than/narrower than’, eschewing the implication of any more specific meaning.)
4. *The criteria for determining when two concepts are equivalent*—in many systems such as SNOMED International and the Clinical Terms v3, there may be several different ways to express the same concept. Human users may be able to recognise that these are essentially the same, but the rules for doing so must be made explicit to be usable by computer.

3.3.2 Differences in granularity

What is required is a language which can handle different levels of detail which supports reliable abstraction along many different axes.

The problem of making information explicit is made more difficult because different amounts of detail, or granularity, are required for different purposes. Two problems must constantly be reconciled. The ability:

1. *To provide sufficient detail* for a particular use
2. *To abstract away irrelevant details* not needed by a particular user

Broadly speaking there are two levels of granularity for clinical systems:

1. The fine grained information required for recording *clinical notes and instructions*
2. The coarse grained information required for *information retrieval, quality assurance, and reporting*.

For clinical care, we must include fine grained detail needed to direct and manage practical care – e.g. a surgical procedure might be described as the amputation of the distal phalanx of the second forefinger. For reporting, a much less detail is required, probably no more than ‘amputation of finger or part of finger’, or perhaps just ‘operation on hand’.

The essence of re-using information in a ‘patient-centred information system’ is to be able to bridge the differences in granularity between these two levels – to be able to abstract from the fine grained information collected for patient care to the coarse grained reporting for management, planning, and remuneration. The process may be obvious to humans, but it must be made explicit for machines.

There are at least five difficulties:

1. *A single fine grained description may be classified in many different ways* under many different coarse grained abstractions for different purposes.
2. *Even the fine grained description may not contain some detail* required for a pre-established set of reporting abstractions. The union of the details required for all potential reporting abstractions may be more than clinical professionals are willing to record while at the same time being less than they require for clinical care.
3. *Not all combinations of fine grained detail can be predicted in advance*,
4. *Not all coarse grained abstractions which may become relevant in the future can be predicted in advance*.
5. *Some coarse grained abstractions may require inference which is not ‘terminological’ because it looks beyond isolated clinical concepts* and requires general inference from the record as a whole: e.g. the classification ‘Cirrhosis without mention of alcohol’ in ICD can only be made with reference to the record as a whole, or at least this episode as a whole, and not simply be looking at a simple clinical statement.

3.3.3 Multiple views and multiple hierarchies: a not so small extension

What is required is a language which can determine the placement of new concepts in a poly-hierarchy based on their definition and descriptions.

Fundamental to the idea of a re-usable terminology is the idea of multiple views for different uses – multiple views in terms of:

1. Axis of classification – anatomical, functional, causal, structural, etc.
2. Granularity and selection of details

Cimino makes two of his desiderata ‘poly-hierarchy’ and ‘concept permanence’ [15], *i.e.* that any concept ought to be capable of having multiple parents and that the each concept should have the same meaning regardless of the parent from which it was reached. ‘Concept permanence’ we take as a consequence of the hypothesis of separability, that in the formal concept representation, a concept can have only one meaning regardless of where it appears. The issue of poly-hierarchy requires more comment.

Early systems were ‘mono-hierarchies’, *i.e.* each concept had only a single parent. Moving to poly-hierarchies – *i.e.* allowing a concept to have more than one parent – may seem a simple change, but in fact it has profound implications.

1. The position in the hierarchy can no longer be expressed by the structure of the code or identifier. In Cimino’s terms, identifiers must be ‘non-semantic’. Hence scheme’s such as ICD’s or Read version 1, 2 (4-digit and 5-digit Read) do not work.

2. 'Inheritance' of default properties can give rise to ambiguity or contradiction. In a simple hierarchy, there is only one route from the top of the hierarchy to any given concept. Hence the simple rule that a concept always inherits the 'nearest' value for a given property works unambiguously. In a poly-hierarchy, there may be multiple routes to concepts, and there is no guarantee that the 'nearest' value will be unique. It has been proven that there is no simple solution to managing default properties in a poly-hierarchy [39].
3. The quantity of information to be maintained is vastly increased. When a new concept is added, all of its potential parents and children must be determined. In a mono-hierarchy each new concept is simply inserted in a sequence from top to bottom. It can have only one parent, and its children must be selected from amongst the children of that one parent. In a poly-hierarchy, a single concept can potentially have indefinitely many parents and indefinitely many children from anywhere in the system. In practice, in a system of any size, it is doubtful that it is possible to maintain a poly-hierarchy consistently by manual means alone.

3.3.4 Scale and the combinatorial explosion

What is required for clinical terminology a system which works in practice to cope with the scale of proposed clinical information systems.

The requirements in the above three sections combine exponentially to produce a problem of enormous scale.

In traditional terminologies, the exponential explosion was managed in two ways:

1. The scope was limited to a single use: single granularity, a single point of view, a mono-hierarchy and often to a single area of medicine.
2. Mono-hierarchies are an efficient way to manage exponentially growing sets of information. A mono-hierarchy, or 'tree', grows only linearly while the total number of items in it grows exponentially. (More precisely, if the tree is 'balanced', the depth grows as the logarithm of the total number of entries)

Unfortunately, neither solution is available if the system of concepts is to be re-usable. If the goal is a concept system which is re-usable, and hence which supports multiple points of view, poly-hierarchies and multiple levels of granularities, are required. From the point of view of scaling, poly-hierarchies are particularly serious: Not only do they sacrifice the logarithmic scaling properties of a mono-hierarchy, but for many calculations, the issue is not just the number of concepts but rather the total numbers of paths through the poly-hierarchy to each concept. The number of paths may be combinatorially greater than the number of concept themselves, *e.g.* if there are five levels and each concept (except the top two levels) has an average of three parents there are 3^3 possible paths from any bottom concept to the top concept.

Ultimately, there is no escape from the combinatorial explosion – it must be paid for either in computation or storage or alternatively limited by restrictions on what can be expressed or inferred. The tradeoffs are the topic of much of computer science and knowledge representation. Much is known, and although the combinatorial explosion cannot be avoided completely, poor choice of techniques can make it catastrophically worse, and good choice of techniques limit it in practical cases.

3.3.5 Open-endedness: Pre-coordination vs Post-coordination

What is required is a system which can be extended smoothly and quickly in response to changing user demands, whether that system is pre-coordinated or post-coordinated,

Not only is medicine big, it is open-ended:

1. *In breadth*, because new information is always being discovered or becoming relevant
2. *In depth*, because finer grained detail is always being discovered or becoming relevant
3. *In complexity*, because new relationships are always being discovered or becoming relevant

No fixed enumerated list of medical concepts can ever be complete. There are two families of solution to this problem:

1. *Pre-coordinated enumerated lists* – *i.e.* lists of concepts with their definitions and descriptions with their classification and relations pre-computed and held in some data structure.
2. *Post-coordinated dynamic classifiers* – *i.e.* software which can allow new concepts to be composed from old concepts and then compute the classification and relations 'on the fly'.

Pre-coordinated solutions fit more easily with conventional computing solutions but require a fast ‘update cycle’ to be satisfactory to users. Post-coordinated solutions can allow users to compose new concepts as needed but require fast, on-line classifiers – a technology which is just becoming available.

3.3.6 Balancing expressiveness and tractability

What is required is a formalism which is sufficiently expressive to capture the information required, computationally tractable for practical cases, and which behaves predictably in the domain.

Thirty years of research in knowledge representation have shown that there are fundamental tradeoffs between the expressivity of the formalism and its computational tractability, *i.e.* between what can be said and what inferences can be computed reliably based on what was said [33, 34, 40]. The early results of Doyle and others [41] with respect to medicine can be caricatured as showing that “Any useful formalism is intractable in the worst case, and any tractable formalism is nearly useless.”

Logic guarantees only that if the premises are true then the conclusions will also be true. For a logical classification system, this means that concepts will only be classified correctly if their definitions and descriptions are correct. If a formalism is not expressive enough, then users tend to misuse it to ‘approximate’ what they want to say – for example by using ‘kind-of’ instead of ‘part-of’. Once the definitions are false, the classification is unpredictable, even if the formalism itself is rigorously correct (‘complete and decidable’). On the other hand, if the formalism is too expressive, then it may not be possible to predictably obtain all results.

Understanding the requirements and tradeoffs and making the best choices for clinical terminologies is a challenging task.

3.3.7 Content, Comprehensibility and Consistency: The Schema or ‘Clinical Ontology’

What is required is an ontology which is both comprehensible and adequate to ensure correct classification and retrieval.

Given any knowledge representation formalism, there are potentially many different ways to use it to represent clinical knowledge. Many choices are constrained by what works in a given language, but often there are several equally effective alternatives amongst which the choice is arbitrary. Because developing a clinical concept representation is too large a task for a single individual, an entire team must agree on the many conventions and stick to them.

The total of the decisions on how to represent the clinical concepts in the chosen language is known as the ‘schema’ or more recently the ‘ontology’ – a word borrowed by knowledge representation from philosophy with perhaps dubious justification[42].

The requirements of developing ontologies in detail are outside the scope of this paper and well discussed elsewhere [25, 36, 43-46].

However, like terminology in general, ontologies combine human ‘analogue’ cognitive requirements with formal logical ‘digital’ requirements. To be successful they must:

1. Be comprehensible and capable of being explained to and used consistently by all members of the team building them.
2. Give the right results when classifying concepts for organising information for retrieval across all applications for which the terminology is to be re-used across all granularities and viewpoints.

These two requirements inevitably conflict. The more re-usable the ontology, the more complex and difficult to build and maintain it becomes. Ultimately, an ontology expressed in a formal knowledge representation language is a computer programme. As the complexity increases the same techniques used to manage complexity in other computer programs may be required, for example defining a ‘higher level language’, or ‘intermediate representation’ in which is more cognitively appropriate for the development team but which can be transformed into the rigorous concept representation automatically [47].

3.3.8 General knowledge, relations, part-whole, and causal relations

What is required is a formalism and ontology capable of dealing with part-whole, causal and other transitive relations in medicine plus the relevant content in the concept representation system.

Medical terminology depends critically on knowledge of how the body fits together anatomically and how pathophysiological processes fit together causally. For example, a concept system cannot support the inference that an injury to the humerus is an injury to the arm unless it has represented within it the fact that the humerus is a part of the arm. Similarly, it cannot infer that the dementia resulting from

AIDS is a viral CNS disorder, unless it has represented within it the fact that AIDS is caused by the HIV virus. Three things are required:

1. A formalism capable of dealing with transitive relations and their interactions with other relations
2. Ontological schemas which are adequate to describe anatomical, causal, and other critical relations
3. The actual content in the descriptions and definitions in the ontology sufficient to make the required inferences.

This implies significantly more content than is present in most traditional 'terminologies'. Furthermore, it requires distinguishing carefully between part-whole and kind-of relations which have often been confused in existing terminologies [47, 48]. Technically also, these requirements present significant problems. The study of part-whole relations is the subject of an entire subfield in linguistics known as 'merology' [49]. It is also the subject of extensive research of formal concept representation [50-52]. However, without this additional information, correct formal classification of medical concepts is impossible.

3.3.9 Conventional and implied information

What is required is a consistent means of coping with conventional and implied knowledge that results in the expected classification of procedures and other clinical concepts.

Many procedures are classified according to implied information about what they achieve or why they are performed. 'Insertion of pins in Femur' would be expected to be classified under 'Fixation procedures' because everyone understands that is what it is for. Should this 'implied' knowledge be part of the concept representation or the representation of clinical pragmatics? Arguments can be made in both directions. If included in the formal concept representation it brings a new set of complications because such statements act as 'rules' whose effects can cascade. If placed in clinical pragmatics, it requires particular care to avoid attributing statements to healthcare professionals which they did not actually make.

4. Consensus, software and systems: the rest of the story and the legacy

Unfortunately, not even solving the three problems of linguistics, pragmatics, and formal representation is sufficient. Clinical terminologies must also fit into larger information and medical record systems – that is, according to our first assumption, what they are for. This involves four further problems.

4.1 Clinical consensus – achieve the appropriate level of agreement

The seventh reason that clinical terminology is hard is that determining and achieving the appropriate level of clinical consensus is hard and requires that the terminology be open ended and allow local tailoring.

Doctors disagree. Nurses disagree. Health care professionals disagree. Achieving clinical consensus on anything is difficult, and consensus on existing terminologies has proved particularly difficult. One of the motivations for the distinctions in this paper and the hypothesis of separability is to minimise the difficulties of achieving consensus. One of the pre-requisites for a re-usable terminology is to establish what level of consensus is appropriate in each area and what areas can be left for local choice.

1. Separating language and concepts allows many disagreements to be circumvented. For example, over the past three decades the use of the word 'neoplastic' and its cognates in various European languages has meant at various times and places either literally 'new growth', *i.e.* 'any cellular proliferation, benign or malignant', or 'malignant proliferation'. Interminable discussions have occurred. However, at no time has there been any question that both underlying concepts were valid, only of which linguistic construct ought to be applied to each. Separating language and concepts allows both to be accommodated while preserving the differences for communication and information sharing. Of the projects which have had to deal most with this problem, GALEN, it takes as one of its mottoes 'Coherence without uniformity' [53].
2. Consensus is not always possible. For example, it is simply a fact that the details of surgical procedures and how they are grouped differ between different European countries. Different diagnoses are accepted in different European countries, and even the same diagnosis may have sufficiently different meaning that the same 'term' must be said to refer to a different concept, *e.g.* 'schizophrenia'. The disciplines of formal representation can help make these differences clear and produce an integrated classification providing a language for expressing which concepts are more specific, more general, or just different from others.

3. Only limited consensus appears necessary for communication with manual records taken on a broad scale. Healthcare professionals manage constantly with minor differences of meaning and even misunderstandings of information written by the colleagues or received from other institutions. Part of clinical practice includes knowing when such mismatches are likely and checking where appropriate. Can we expect more of computer-based systems?
4. The degree of consensus required depends on local factors. Consistent quality assurance, decision support, planning, and perhaps remuneration, depend on varying degrees of consensus. Local tailoring and adaptation is therefore essential, contributing to the view that medical terminology must be open ended (See 3.3.5).

However, even given the above, achieving sufficient clinical consensus even on broad issues requires effort.

4.2 The legacy of existing classification and coding systems

The eighth reason clinical terminology is hard is because the structural idiosyncrasies of existing coding and classification systems must be addressed.

The structure of existing classification systems – ICD, CPT, OPCS, etc.– represent existing structures to which any new system for use in software must map and from which it must be able to take information. These systems will be essential for reporting and other functions for the foreseeable future. They have complex structure in their own right which must be represented, including awkward constructs which have no direct equivalent in a formal concept representation, linguistic, or pragmatic representations such as ‘not elsewhere classified’, ‘other’, etc.

4.3 Meeting the needs of electronic patient records

The ninth reason clinical terminology is hard is that it must be coordinated and coherent with models of the electronic patient record

The fundamental assumption of this paper is that the purpose of clinical terminology is to support clinical software in general and Electronic Patient Records in particular. The structure of the terminology must be coordinated with that of the medical record. For example, there is often a choice of whether the same information should be recorded in the structure of the medical record or in the system of concepts – for example the extent of surgery might be either part of the ‘concept’ recorded in a single field in the patient record or it might be recorded as a separate field. Worse, an identifier for a ‘local mastectomy’ might be recorded in the field for surgical procedure and an identifier for ‘radical’ in the field for extent. The information should at a minimum be consistent, and ideally, the same information should be represented only once, either in the medical record or in the terminology. Furthermore, information about the context of the events as well as the events themselves must be expressed. Hence there are mutual constraints between the Electronic Patient Record model and the terminology. Rossi Mori has discussed this problem in detail [54, 55]. However, currently the interface between formal concept and the developing models for medical records remains less well defined than required for software systems and standardisation.

4.4 Change management and persistence in the medical record

The tenth reason that clinical terminology is hard is that change must be managed, and it must be managed without corrupting information already recorded in medical records.

Change is inevitable, regardless of the type of system used, whether pre- or post-coordinated, whatever the representation language used. Aspects of change management for clinical terminologies has been extensively discussed by other authors, *e.g.* Oliver[56], Campbell [57], and Cimino [58].

However, it is worth noting that within the framework presented here based on the ‘hypothesis of separability’, there are at least three sorts of potential changes to be dealt with:

1. Changes of language with respect to the underlying concepts
2. Changes in clinical practice
3. Changes to the underlying system of concepts itself

For example, when AIDS was first described, there was nothing that could be said with certainty about its aetiology, and even its definition was fuzzy. Today we can at least agree that it is caused by HIV and probably that it can be defined as ‘the clinical syndrome caused by HIV infection’. Note that three things have must happen to implement such a change:

1. A new virus, HIV, must be added to the list of viruses in the concept model.

2. What is relevant to different sorts of patients, both those with AIDS and those with other conditions, must be changed; what may be assumed in context may also have changed.
3. The language construct 'AIDS' must be moved from the earlier ill defined concept to the newer more precisely defined concept.

Clearly the earlier ill defined concept and the newer more precisely defined concept of AIDS are not the same, and any faithful record system must maintain the difference. On the other hand, any search for patients who had AIDS in 1980 should retrieve patients who were given that diagnosis then, even though it corresponds to the earlier ill-understood concept. Both the old and new concept representations are still valid descriptions of clinical concepts, although no one would use the 1980 version in 1999. As far as the author is aware, no one has proposed or implemented a complete solution to this problem even in a mono-lingual terminology, let alone in a multi-lingual one.

At a minimum being sure that all of the information in the medical record is retained faithfully requires that both precise version control and permanent storage in the medical record of both the language and the formal representation. The hypothesis of separability does not imply that one component is more fundamental than the other or that one can be substituted for the other, only that they be processed separately.

5. Discussion:

5.1 Relation to other work

Although written from a different perspective, this paper has drawn heavily on Cimino's paper on 'Desiderata for Controlled Medical Vocabularies...' [15] and Rossi Mori's approach to generations of terminologies [59]. Rossi Mori's second and third generation systems correspond roughly, though not precisely, to pre- and post-coordinated systems as described in Section 3.3.5.

Table 1 gives a rough correspondence of Cimino's desiderata and the issues discussed in this paper. As can be seen, there is broad agreement on substance, but the issues are grouped very differently. Many of Cimino's desiderata follow from what this paper terms the 'hypothesis of separability'. Others, such as 'concept permanence' (by which Cimino means that a concept means the same thing however it is reached in a poly-hierarchy), follow from the view of formal representation given here, that in a formal concept representation system, concepts are, by definition, unique and constant. What Cimino has listed as content and coverage we have taken as one of the important sources of the scaling problems that follow from the aspiration of re-use.

5.2 Separations and interdependencies: Pragmatics and implementation

This paper proposes that a medical terminology system must solve simultaneously problems from at least three primary disciplines:

1. Clinical linguistics
2. Clinical pragmatics
3. Formal concept representation

To which must be added at least three additional functions

1. Interfaces to existing classification and coding systems
2. Interfacing to medical records and wider clinical information systems
3. Change management and persistence

Assuming separate components for each, the degree of overlap and mutual constraints between the components is still not well understood. Traditional development of terminologies has focused on information retrieval as their primary task, and so fall closest to concept representation, however informal. By contrast, most have been at best 'semi detached' from clinical pragmatics – *i.e.* 'implementation', data-entry, and user interfaces. In the meantime, developers of messaging systems have, by and large, developed special purpose vocabularies or 'value sets' for their own purposes, though these are increasingly being incorporated into broader terminologies.

In short, for three of the four primary tasks cited in the introduction - entering information, retrieving information, sharing information – there are terminologies suited to each but none to all. Furthermore, there are systems taking each as their primary criteria for evaluation but few which strive to serve, or even admit the validity of, all four needs. (Exceptions include our own work on PEN&PAD [27, 60] based on GALEN and Lussier's work based on SNOMED [61]).

5.3 What can be done? What are the priorities?

Faced with a hard problem there are, roughly speaking, four possible approaches:

1. Simplify the problem to achieve the highest priorities
2. Apply more effort
3. Combine efforts
4. Get better tools

Given that for many medical systems the priority is defining messages for communication, the first solution is potentially appealing – define limited vocabularies and ‘value sets’ for limited purposes and declare everything else to be out of scope. This appears to be an approach taken by message developers such as HL7 and LOINC with considerable success. There are clearly potential long term costs in integration, but the short term advantages may well outweigh them. There may be other areas where similar prioritisation is the best way forward, and it may be that some of the aspirations for ‘patient centred systems’ are best deferred until more urgent priorities have been dealt with.

Applying more effort has perhaps been most typical of the UK Clinical Terms project (Read Codes) version 2, where forty groups of clinicians laboured for several years to produce definitive sets of codes and rubrics. In the end there were between 250,000 and 500,000 terms yet still insufficient clinical detail. The change in direction, several years after the projected delivery date, to a more structured semi-formal approach for ‘version 3’ [62] speaks for itself of the difficulties encountered. Given the scale of the project, it can be taken as strong evidence that just applying more effort does not work.

On the other hand, combining efforts along with an investment in specific types of tools as the UMLS has done has produced a range of useful results. The availability of a cross reference, even with imperfections, amongst the various existing coding systems does not overcome their intrinsic limitations, but it often make it possible to compare results and link information, particularly external knowledge sources and bibliographic information [63-65] and, in some cases in medical records and problem statements [66].

Getting better tools is the approach, in different ways, of the SNOMED-RT project [67] and GALEN [68-70]. To a lesser extent, the UK Clinical Terms, version 3 also applies the ‘better tools’ philosophy, but development of the clinical terms remains, fundamentally a manual process. Comparison of the results of the Clinical Terms and GALEN’s techniques suggest that even the most systematic and scrupulously careful manual attempts to produce a multi-axial compositional system will be incomplete and inconsistent in significant ways. However, they also show the importance of manual quality checks for any more formal system such as GALEN [71]. Manual effort alone cannot guarantee consistency; tools cannot guarantee clinical validity. The results of SNOMED-RT [13] are still pending.

Both GALEN and SNOMED-RT are significantly hampered in different ways by deficiencies in the knowledge representation languages (description logics) used. One of the hopeful signs for future systems is that the technology of expressive representation languages is improving rapidly so that many of these limitations may be overcome in future [72, 73].

It is less clear whether progress is being made in terms of the content and ontology. The Digital Anatomist project has produced and tested a rigorous ontology for anatomy [43]. GALEN’s ontology has apparently stood up to the challenge of developing surgical classifications for a variety of centres, as has its approach to building the ontology using an Intermediate Representation [47, 74]. However, convincing evaluation and a more generic theory of ontologies analogous to the theory for formalisms for knowledge representation is still lacking.

5.4 Terminology as Software: Towards an integrated approach

The first assumption of this paper is that the purpose of clinical terminology is to support clinical software in the context of aspirations for ‘patient based systems’ Many issues have been identified that make developing such a clinical terminology a ‘hard problem’. However, there are three relatively simple fundamental conclusions.

Firstly, that a terminology to support ‘patient centred’ information systems in which all, or most, information is derived automatically from the Electronic Patient Record is of vastly greater scale than traditional terminologies such as ICD, The Clinical Terms (read codes) versions 1 and 2, and even SNOMED International.

Secondly, the resulting scale exceeds what can be managed manually with the rigour required by software, but building appropriate rigorous representations on the necessary scale is, in itself, a hard problem.

Thirdly, there is a fundamental conflict between the needs of software and the needs of human users. In addressing this conflict, the need for scalable 'clinical pragmatics' – practical data entry, presentation and retrieval for clinical tasks – must be taken into account. Too often clinical pragmatics has been either neglected or explicitly excluded as 'implementation'. Addressing the need to support clinical pragmatics means that validation of clinical terminologies must include validation in use implemented in software. Not to do so is to validate terminologies for use '*in vivo*' on the basis solely of '*in vitro*' experiments.

The final question in the introduction was "How will we know if we have succeeded?". We will know we have succeeded when clinical terminologies in software are used and re-used, and when multiple *independently developed* medical records, decision support, and clinical information retrieval systems sharing the same information using the same terminology are in routine use.

References

1. Sittig, DF, *Grand challenges in medical informatics*. Journal of the American Medical Informatics Association, 1994. **1**: pp. 412-413.
2. CEN TC251. *CEN TC251 Home Page*. 1999: CEN TC251: <http://www.centc251.org/>
3. ISO TC215. *Home Page*. 1999: ISO TC215: <http://www.imc.exec.nhs.uk/isotc215/>
4. ISO TC215 WG3. *Working Group on Terminology, Home Page*. 1999: <http://www.imc.exec.nhs.uk/isotc215/>
5. HL7. *Health Informatics Standards - HL7*. 1999: Duke University: <http://www.mcis.duke.edu/standards/HL7/hl7.htm>
6. DICOM. *DICOM Standards Links*. 1999: <http://www.mcis.duke.edu/standards/DICOM/dicom.htm>
7. Solbrig, H, *Final submission to the CorbaMED Request for Proposals on Lexical Query Services (CorbaLex)*, . 1998, OMG:
8. United States Senate, *Health Insurance Portability and Accountability Act, S1028*. 1996.
9. Rogers, R and Reardon, J, *Barriers to a Global Information Society for Health: Recommendations for International Action, Report from the project G8-Enable*. 1999, Oxford: IOS Press.
10. Walther, E, Eriksson, H, and Musen, MA. *Plug-and-Play: Construction of task-specific expert-system shells using sharable context ontologies*. in *AAAI Workshop on Knowledge Representation Aspects of Knowledge Acquisition*. 1992. San Jose CA: pp. 191-198.
11. Musen, M, Tu, S, Das, A, and Shahar, Y, *EON: A component based architecture for automation of protocol-directed therapy*. Journal of the American Medical Informatics Association, 1996. **3**: pp. 367-383.
12. Das, SK, Fox, J, and Krause, P. *A unified framework for hypothetical and practical reasoning (1): theoretical foundations*. in *International Conference on Formal and Applied Practical Reasoning*. 1996: Berlin: Springer Verlag.
13. Campbell, K, *Scalable methodologies for distributed development of logic-based convergent medical terminology*. *Methods of Information in Medicine*, 1998. **37**(426-439).
14. Spackman, KA, Campbell, KE, and Côté, RA, *SNOMED-RT: A reference Terminology for Health Care*. Journal of the American Medical Informatics Association (JAMIA), 1997. **(Symposium special issue)**: pp. 640-644.
15. Cimino, J, *Desiderata for controlled medical vocabularies in the twenty-first century*. *Methods of Information in Medicine*, 1998. **37**(4-5): pp. 394-403.
16. ISO, *ISO 1087: Terminology - Vocabulary*, . 1990, ISO:
17. NHS National Health Service Executive, *Information for Health: An information strategy for the modern NHS 1998-2005*, . 1998, UK National Health Service Executive: <http://www.imt4nhs.exec.nhs.uk/>.
18. Dick, RS and Steen, EB, eds. *The Computer-Based Patient Record: An essential technology for health care*. . 1991, National Academy Press: Washington, DC.
19. Norman, DA, *The Invisible Computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*. 1998, Cambridge Mass: MIT Press.
20. Sager, N, Lyman, MS, Bucknall, C, Nhan, NT, and Tick, LJ, *Natural Language Processing and the Representation of Clinical Data*. Journal of the American Medical Informatics Society, 1994. **1**(1): pp. 142-160.
21. Baud, RH, Rodrigues, J-M, Wagner, JC, Rassinoux, A-M, Lovis, C, Rush, P, Trombert-Paviot, B, and Scherrer, J-R, *Validation of concept representation using natural language generation*. Journal of the American Medical Informatics Association, 1997. **(Fall Symposium Supplement)**: pp. 841.
22. Lovis, C, Baud, R, Michel, P-A, and Scherre, J-R, *Morphosemantic decomposition and semantic representation to allow fast and efficient natural language recognition*. Journal of the American Medical Informatics Association, 1997. **(Fall Symposium Supplement)**: pp. 873.

23. Rasinoux, A-M, Juge, C, Michel, P-A, Baud, R, Lemaitre, D, Jean, F-C, Degoulet, P, and Scherrer, J-R. *Analysis of medical jargon: The RECIT system*. in *Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95)*. 1995. Pavia, Italy: Springer: pp. 42-42.
24. Wagner, J, Rasinoux, A-M, Baud, R, and Scherrer, J-R. *Generating noun phrases from a medical knowledge representation*. in *Twelfth International Congress of the European Federation for Medical Informatics, MIE-94*. 1994. Lisbon, Portugal: pp. 218-223.
25. Zweigenbaum, P, Bachimont, B, Bouaud, J, and Charlet, J, *Issues in the structuring and acquisition of an ontology for medical language understanding*. *Methods of Information in Medicine*, 1995. **34**(1/2): pp. 15-24.
26. Hahn, U, Romacker, M, and Schulz, S, *How knowledge drives understanding - matching medical ontologies with the needs of medical language processing*. *Artificial Intelligence in Medicine*, 1999. **15**(1): pp. 25-52.
27. Nowlan, WA, *Clinical workstation: Identifying clinical requirements and understanding clinical information*. *International Journal of Bio-Medical Computing*, 1994. **34**: pp. 85-94.
28. Kirby, J, Cope, N, Souza, Ad, Fowler, H, and Gain, R. *The PEN&PAD Data Entry System*. in *Medical Informatics Europe (MIE-96)*. 1996. Copenhagen: IOS Press: pp. 430-434.
29. Moorman, P, van Ginneken, AMvG, Siersema, PD, van der Lei, J, and van Bommel, JH, *Evaluation of reporting based on descriptive knowledge*. *JAMIA*, 1995. **2**(6): pp. 365-373.
30. Kay, S and Purves, IN, *Medical records and other stories: a narratological framework*. *Methods of Information in Medicine*, 1996. **35**(2): pp. 72-87.
31. Simon, HA, *The Sciences of the Artificial*. 2 ed. 1969, 1981: MIT Press. 247.
32. *British National Formulary*. 1998: British Pharmaceutical Society.
33. Borgida, A, *Description logics in data management*. *IEEE Transactions on Knowledge and Data Engineering*, 1995. **7**(5): pp. 671-682.
34. MacGregor, R, *The evolving technology of classification-based knowledge representation systems*, in *Principles of Semantic Networks: Explorations in the representation of knowledge*, J. Sowa, Editor. 1991, Morgan Kaufmann: San Mateo, CA. p. 385-400.
35. Sowa, J, *Conceptual Structures: Knowledge Representation in Mind and Machine*. 1985, New York: John Wiley & Sons.
36. Lenat, DB and Guha, RV, *Building Large Knowledge-Based Systems: Representation and inference in the Cyc Project*. 1989, Reading, MA: Addison-Wesley. 372.
37. Fikes, R, Cutkosky, M, Gruber, T, and Baalen, jV, *Knowledge sharing technology -- project overview*, . 1991, Stanford University, Knowledge Sharing Laboratory:
38. Bateman, JA, *Upper modelling: a general organization of knowledge for natural language processing*, . 1989, USC/Information Sciences Institute:
39. Touretzky, D, *The Mathematics of Inheritance Systems*. 1986, Los Altos, CA: Morgan Kaufmann.
40. Brachman, R and Levesque, H. *The tractability of subsumption in frame-based description languages*. in *AAAI-84*. 1984: Morgan Kaufman: pp. 34-37.
41. Doyle, J and Patil, R, *Two theses of knowledge representation: Language restrictions, taxonomic classification and the utility of representation services*. *Artificial Intelligence*, 1991. **48**: pp. 261-297.
42. Guarino, N and Giarretta, P, *Ontologies and knowledge bases: towards a terminological clarification*, in *Towards Very Large Knowledge Bases*, N. Mars, Editor. 1995, IOS Press: Amsterdam. p. 25-32.
43. Rosse, C, Shapiro, IG, and Brinkley, JF, *The Digital Anatomist foundational model: Principles for defining and structuring its concept domain*. *Journal of the American Medical Informatics Association*, 1998(1998 Fall Symposium Special issue).
44. Rector, A. *Coordinating taxonomies: Key to re-usable concept representations*. in *Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95)*. 1995. Pavia, Italy: Springer: pp. 17-28.
45. Rogers, J and Rector, A. *The GALEN ontology*. in *Medical Informatics Europe (MIE 96)*. 1996. Copenhagen: IOS Press: pp. 174-178.

46. Farquhar, A, Fikes, R, Pratt, W, and Rice, J, *Collaborative Ontology Construction for Information Integration*, . 1996, Knowledge Systems Laboratory, Stanford University:
47. Rector, AL, *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): pp. 229-242.
48. Rector, A, *Thesauri and formal classifications: Terminologies for people and machines*. Methods of Information in Medicine, 1998. **37**(4-5): pp. 501-509.
49. Cruse, DA, *Lexical Semantics*. 1986, Cambridge, England: Cambridge University Press.
50. Padgham, L and Lambrix, P. *A Framework for Part-of Hierarchies in Terminological Logics*. in *KR-94*. 1994: pp. 485-96.
51. Odell, JJ, *Six different kinds of composition*. Journal of Object Oriented Programming, 1994. **5**(8): pp. 10-15.
52. Winston, M, Chaffin, R, and Hermann, D, *A taxonomy of part-whole relations*. Cognitive Science, 1987. **11**: pp. 417-444.
53. Rector, A, Rossi Mori, A, Consorti, F, and Zanstra, P, *Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation*. International Journal of Medical Informatics, 1998. **48**(1-3): pp. 71-84.
54. Rossi Mori, A, Galeazzi, E, Consorti, F, and Bidgood, WD, *Conceptual Schemata for Terminology: A continuum from headings to values in patient records and messages*. Journal of the American Medical Informatics Association, 1997. (**Fall Symposium Special Issue**): pp. 650-654.
55. Rossi Mori, A and Consorti, F, *Integration of clinical information across patient records: a comparison of mechanisms used to enforce semantic coherence*. IEEE Transactions on Information Technology in BioMedicine, 1999. **4**(2): pp. 243-253.
56. Oliver, DE, Shahar, Y, Shortliffe, EH, and Musen, MA, *Representing change in controlled medical vocabularies*. Artificial Intelligence in Medicine, 1999. **15**(1): pp. 53-76.
57. Campbell, K, Cohn, S, Chute, C, Rennels, G, and Shortliffe, E. *Gálapagos: Computer-based support for evolution of a convergent medical terminology*. in *AMIA Fall Symposium*. 1996. Washington DC: Hanley and Belfus, Inc: pp. 269-273.
58. Cimino, J and Clayton, PD. *Coping with changing controlled vocabularies*. in *Eighteenth Annual Symposium on Computer Applications in Medical Care*. 1994. Washington, DC: Hanley & Belfus, Inc, Philadelphia PA: pp. 135-139.
59. Rossi Mori, A, Consorti, F, Galeazzi, E, and Merialdo, P. *A second generation of terminological systems is coming*. in *Medical Informatics Europe '97*. 1997. Porto Carras, Greece: IOS Press: pp. 436-440.
60. Kirby, J and Rector, AL. *The PEN&PAD Data Entry System: From prototype to practical system*. in *AMIA Fall Symposium*. 1996. Washington DC: Hanley and Belfus, Inc: pp. 709-713.
61. Lussier, YA, Maksud, M, Desruisseaux, B, Yale, P-P, and St-Arneault, R. *PureMD: A computerized patient record software for direct data entry by physicians using keyboard-free pen-based portable computer*. in *Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC-92)*. 1992. Baltimore, MD: McGraw-Hill: pp. 261-263.
62. Brown, P, O'Neil, M, and Price, C, *Semantic definition of disorders in Version 3 of the Read Codes*. Methods of Information in Medicine, 1998. **37**: pp. 415-419.
63. Shortliffe, E, Barnett, G, Cimino, J, Greenes, R, Huff, S, and Patel, V, *Collaborative medical informatics research using the Internet and the World Wide Web*. Journal of the American Medical Informatics Association, 1996. **1996 Symposium Supplement**: pp. 125-129.
64. Pratt, W, *Dynamic organisation of search results using the UMLS*. Journal of the American Medical Informatics Association, 1997(1997 Fall Symposium Special Issue): pp. 480-484.
65. Detmer, WM, Barnett, GO, and Hersh, WR, *MedWeaver: Integrating decision support, literature searching and web exploration using the UMLS Metathesaurus*. Journal of the American Medical Informatics Association, 1997(1997 Fall Symposium Special Issue): pp. 490-494.

66. Goldberg, HS, Hsu, C, Law, V, and Safran, C, *Validation of clinical problems using a UMLS based semantic parser*. Journal of the American Medical Informatics Association, 1998(1998 Fall Symposium Special Issue): pp. 805-809.
67. Spackman, KA and Campbell, KE, *Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies*. Journal of the American Medical Informatics Association, 1998(1998 Fall Symposium special issue): pp. 740-744.
68. GALEN, *Home Page URL: <http://www.cs.man.ac.uk/mig/galen>*, . 1997:
69. Rector, AL, Baud, R, Ceusters, W, Claassen, W, Rodrigues, J-M, Mori, JRR, Haring, Evd, Solomon, WD, and Zanstra, P, *A comprehensive approach to developing and integrating multilingual classifications: GALEN's Classification Workbench*. Journal of the American Medical Informatics Association, 1998. **Fall Symposium Special Issue**: pp. 1115.
70. Rector, A, Rossi Mori, A, Consorti, F, and Zanstra, P. *Making sound re-usable terminology practical: The GALEN approach*. in *Towards and Electronic Health Care Record, Europe*. 1998. London: Medical Records Institute, Newton Mass: pp. 298-305.
71. Rogers, JE, Price, C, Rector, AL, Solomon, WD, and Smejko, N, *Validating clinical terminology structures: Integration and cross-validation of Read Thesaurus and GALEN*. Journal of the American Medical Informatics Association, 1998. **Fall Symposium Special Issue**(845-849): pp. in press.
72. Horrocks, I. *The CAMELOT Web Page*. 1998: <http://www.cs.man.ac.uk/~horrocks/Camelot/>
73. Horrocks, I. *Using an expressive description logic: FaCT or Fiction*. in *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference on Knowledge Representation (KR 98)*. 1998. San Francisco, CA: Morgan Kaufmann: pp. (in press).
74. Rogers, J, Solomon, D, Rector, A, and Zanstra, P. *From rubrics to dissections to GRAIL to classifications*. in *Medical Informatics Europe (MIE-97)*. 1997. Thesalonika, Greece: IOS Press: pp. 241-245.

Table 1: Comparison of Cimino's Desiderata and issues in this paper

Cimino's Desiderata	Relevant Issues in this paper
Concept Orientation	Separation of language and concepts (3.1)
Content/Coverage	Scale and diversity of use (2.1)
Comprehensiveness	Scale and diversity of use (2.1)
Non-ambiguity	Separation of language and concepts (3.1) Formal Concept Representation (3.3)
Poly-hierarchy	Multiple views and multiple hierarchies (3.3.3)
Formal Definitions	Formal Concept Representation (3.3)
Multiple consistent views	Multiple views and multiple hierarchies (3.3.3)
Multiple granularities	Differences in granularity (3.3.2)
Nonsemantic identifiers	Multiple views and multiple hierarchies (3.3.3)
Recognise redundancy	Criteria for classifying concepts Criteria for determining when two concepts are equivalent (3.3.1 items 3-4)
Evolve gracefully	Change Management (4.4)
Concept permanence	Formal Concept Representation (3.3) Multiple views and multiple hierarchies (3.3.3) Separation of language and concepts (3.1)
Reject 'not elsewhere classified'	The legacy of existing classifications and coding systems (4.2)
Representing Context	Meeting the needs of electronic patient records (4.3)