Ontological Issues in using a Description Logic to Represent Medical Concepts: Experience from GALEN

Alan Rector and Jeremy Rogers

Medical Informatics Group
Department of Computer Science
University of Manchester
&
The *Open*GALEN Foundation

Abstract

Concept models, or 'ontologies' represented in description logics are becoming central to many efforts in re-usable medical terminologies. Developers of any ontology must make a series of choices concerning how concepts are represented. These choices are made more difficult by the limitations of the logical formalisms in which ontologies are represented and the complexity of the pragmatics of medical usage, as well as by the sheer size and scope of medicine. Principles for ontology structure to guide developers are still poorly. This paper articulates the principles arrived at from the experience of the GALEN programme and illustrates them with an overview of the its ontology, the GALEN Common Reference Model. The complete Common Reference Model and associated material is available from the *Open*GALEN foundation at www.OpenGALEN.org

Citation: Appeared at IMIA WG6 Workshop: Terminology and Natural Language in Medicine. Phoenix Arixona, Nov 1999.

Two appear as a two part paper: Rector, AL and Rogers JE (2002) Ontological issues in GALEN: Part I Principles and Part II High Level Schemata in *Methods of Information in Medicine.*

1. Introduction

GALEN seeks to provide re-usable terminology resources for clinical systems. The heart of GALEN is the use of an ontology, the Common Reference Model, formulated in a specialised description logic, GRAIL [1]. There are several other efforts to use description logics or closely related formalisms in medical terminology including SNOMED-RT [2-4], the pioneering work of the 'Canon group' [5-7], and more language oriented work, ***e.g.*** [8, 9], plus less formal methods of representing the semantics of terminologies used in the NHS Clinical Terms (Read Codes) [10]. Ontologies based on related formalisms play a key role in attempts to develop re-usable problem solving methods by by Musen and his colleagues in [11-13]. Outside of medicine, description logics are being increasingly used for knowledge representation, indexing, and data management in a wide variety of fields, ***e.g.*** [14-16], and ontologies form the foundation for major projects in knowledge representation such as Cyc [17] and the knowledge sharing effort [18]. Increasingly such methods are being advocated for encoding 'meta knowledge' for the Web. [19, 20]

Any attempt to represent clinical concepts using a logic-based formalism must address certain issues and make certain choices. These choices are embodied in the high level schemas of the resulting logical model or ontology. Because many of the issues result from conflicts between users' intuition and the behaviour of logical systems, many features of the resulting ontological schemas are difficult to understand when first seen.

This paper presents a unified approach to the GALEN ontological schemas. Partial presentations have appeared in [21, 22]. It is supplemented with additional material on the *Open*GALEN web site www.OpenGALEN.org, its documentation, http://www.opengalen.org/open/crm/CRM.html, and in the appendices. It is intended to provide an accessible introduction to the GALEN ontology and its rationale. However, it concentrate on general issues in which any clinical ontology must face and hopes to provide sufficient detail to open further debate on alternative approaches to these issues.

This paper assumes a general familiarity with the ideas of formal classification and compositional concept representation. More detailed information and introductions can be found in [1, 23-25] and on the *Open*GALEN web site.

1.1 Kinds of knowledge – Modules of a terminology system

A key feature of the GALEN approach is that it divides the problem of clinical knowledge representation and terminology into distinct parts each served by different software modules in the GALEN implementations as listed in Table 1. This paper concentrates on just one aspect of the GALEN approach to terminology, the concept representation system – resource a). For this paper the most important function of Table 1 is to say what the concept representation, or ontology, is not. It is a key element of a terminological system, but other resources are required for language, coding, user interfaces, and many other reasoning tasks.

High level issues related to the overall architecture and desiderata for terminologies [26] can be found in [23]. The GRAIL formalism is described in [1] the perspectives and intermediate representations in [27, 28], the linguistic resources in [29-33] , and the terminology server in [34, 35]. The use of the GALEN model for indexing is discussed in connection with drug information in [36].

Table 1: GALEN Resources/Modules

a)	*The concept representation, or ontology, schema and model expressed in a description logic, GRAIL*
b)	The linguistic resources for presenting the model
c)	The mappings to and from the concept model and other representations
d)	Perspectives, views and intermediate representations of the model which adapt it to particular purposes
e)	Indexes to other knowledge based on the model
f)	Non-terminological computational or other reasoning mechanisms, ***e.g.*** unit conversion
g)	The terminology server and its API which make all of the other parts available as a coherent whole to applications and users.

Table 2: Elements of an Ontology based on a Description Logic

a)	A hierarchy of elementary categories .
b)	A hierarchy of semantic links for connecting the elementary categories. In GALEN the semantic links are called 'attributes' (corresponding to 'roles' in description logics or 'relations' in typical object oriented modelling'). *Note that in description logics and related formalisms, 'is-kind-of' ('subsumption') is NOT a semantic link but a logical inference of the form "All Bs are As" (see [16, 24]).*
c)	A set of definitions of composite concepts in terms of the elementary concepts: ***e.g.*** of *'Fracture of Femur' = Fracture <u>which</u> hasLocation Femur, 'Foot bone' = Bone <u>which</u> isStructuralComponentOf Foot*, etc.
d)	A set of further facts or 'axioms' about the concepts. These axioms are always of the logical form "All X haveLinkTo some Y" ***e.g.*** that "All feet are a division of some LowerExtremity", "All femurs are a structural component of some thigh". In GALEN such axioms are expressed as 'necessary statements' – see [24]
e)	A set of constraints on what concepts may be linked to each other by specific semantic links. In GRAIL these constraints take the form of 'sanctions' – see [24].

1.2 Goals and Criteria

The key elements of the ontology itself are given in Table 2

The overall aim of the terminology resources is support clinical information systems – to allow information to be recorded in electronic patient records, abstracted from them, re-organised to provide clearer views of an individual patients, aggregated for management, research, and administration, and linked to knowledge resources – decision support, bibliographic, and general web-based information systems.

Key goals for the ontology can be formulated in terms of:

a) *Expressiveness* – the ability to represent formally the concepts clinicians use: symptoms, diseases, procedures, etc.
b) *Abstraction* – the ability to define generic categories of those concepts, describe the relations between them, and to organise them reliably.

GALEN achieves expressiveness by providing a compositional representation of concept representations. It provides abstraction by supporting formal logical classifications of the concepts represented. GALEN achieves scalability and maintainability by providing formal criteria and algorithms for determining whether compositions are consistent and performing and classification automatically.

GALEN's original idealised goal was an ontology which could express 'all and only' what it was medically sensible. It was recognised from the start that this ideal was unobtainable and that the 'all' would have to take precedence over the 'only' because:

a) There are well known trade-offs between expressiveness and computational tractability in formal systems [37, 38].
b) Reality is fractal – no matter how much detail is represented in the model, it is always possible to represent more
c) Constraints on what is 'sensible' are slippery and difficult to formulate loosely enough to allow all sensible statements but strictly enough to exclude all nonsense

The practical goal has been to have an ontology which is sufficiently expressive to capture the statements and abstractions used by clinicians and classify them correctly while rejecting patent nonsense.

2. Principled development of clinical ontologies

2.1 Basic Prinicples

2.1.1 'Logical approximations', labels of concept representations

Any logical model for knowledge representation is at best an approximation of the concepts represented as used in human language and thought. A 'Logical approximation' may seem an oxymoron, but given the fractal nature of reality and the known trade-offs between expressiveness and tractability, this is the most accurate description of the ontologies we build. Logical models of any kind behave very differently from language or 'units of thought'. The flexible fluid dependence on context typical of thought and language eludes logical representations. Two important reasons that effect ontologies are:

Since any ontology is a approximation, the labels attached to concept representations internally in the ontology are at best an approximations. Arguments such as "Is the hand still a division of the upper extremity if its been amputated?" or "Is there a difference between an 'act' and a 'deed'?" are largely futile. When arguments over the labelling of concept representations occur, GALEN asks into two questions:

a) Does the representation represent ***some*** concept which most agree is useful and valid even if they cannot agree on what to call it?
b) Is the label seriously misleading? Ambiguous? Does it mean different things to different groups?

On the whole GALEN has found that non-understanding, puzzlement, and linguistic awkwardness are better than misunderstanding or, worse, multiple understandings, at least for the internal form of the ontology. Hence internal labels often appear awkward or strange, ***e.g.*** *IntrinsicallyPathologicalPhenomenon*.

2.1.2 Principled separation of categories and instances

The GALEN concept model is about ***categories*** (also known as 'types', or 'classes'). Categories can be abstract, such as 'Phenomenon' or 'Disease', general such as 'Diabetes' or very specific such as 'aspirin' or 'foot'. Categories can be specialised – ***e.g.*** "aspirin" can be specialised to "children's aspirin" "foot" can be specialised to "left foot", "deformed foot", "deformed left foot", etc.

Statements about the real world in medical records represent ***instances*** ('individuals') of these categories. The GALEN Common Reference Model itself contains no instances. Instances can only be described but not specialised. It makes no sense to say "a kind of this tablet of Aspirin" or "a kind of Alan's foot". This view corresponds closely to the view that the 'terminology fills the boxes in the information model' propounded in [39, 40] and in recent HL7 discussion documents. It leads to a clean separation of the information model of the medical record and the model of the concept representations it holds.

Many ontologies identify instances as simply being concepts specialised to the level of detail required for a particular application – often determined by that is required for the entries in tables in a database (see Brachman's 'Living with Classic' [41]). Unfortunately, such an approach is fatal for aspirations to re-use, since, as Brachman so elegantly demonstrates, the appropriate level of detail for one application will almost certainly be different from that for another.

2.1.3 Natural kinds – what should be elementary

While GALEN takes all concepts in the concept model or terminology to be categories or types rather than instances, there is still an important issue as to what should be represented as 'elementary' and what by defined composite. For GALEN this is roughly the equivalent of the issue of what should be an instance as discussed by Brachman in [41].

There are two issues:

a) Whether it is *possible* to define the concept. A definition must give necessary ***and sufficient*** criteria for recognising that concept. Many important concepts defy definition by sufficient criteria. Such concepts are termed 'natural and include most simple notions such as "leg", or "tree", "process", "flow", etc.

b) Whether it is *useful* to define the concept with respect to the needs of the applications to be supported by the model. Some concepts are simply not worth the trouble to define even if definition might be possible.

Language can be poor guide to making these decisions. In general, single words are likely to name natural kinds and phrases to give definitions. However, medical language is full of phrases such as 'black bird' which appear to define a composite category but are actually idiomatic names for natural kinds such as 'blackbird'. "Hepatic Artery" names a specific anatomical structure; it does not mean 'Artery which serves the liver' or 'Artery in the Liver', both of which include a number of arteries of which the Hepatic Artery is only one. Hence, in GALEN all body parts are elementary concepts down to the level of laterality and position – ***i.e.*** down to the level of 'finger' or 'toe'.

Natural kinds can also occur at a more abstract level. One might be tempted to define "heart valve" equivalently to "valve in the heart", and "valve" as a "structure which functions as a valve". But this combination results in the "foramen ovale" being classified as a "heart valve", since it undoubtedly functions as a valve. Similarly, one might be tempted to define "endocrine surgery" equivalently to "surgery on an endocrine organ", but this includes "correction of torsion of the testes" and "oösalpingectomy". In both cases it is better to treat the abstractions as natural kinds and then describe them as "function as a valve" and "being on an endocrine organ" respectively. These properties are *necessary,* but to treat them as ***sufficient*** results in over-generalisation with respect to common clinical usage.

2.1.4 Explicitness and orthogonal taxonomies

Potentially, it should be possible to classify any concept on any of its properties independently – ***i.e.*** it should be possible to re-arrange the ontology along any axis. To do so, those properties must be represented explicitly and independently, even at the cost of apparent redundancy. For example, people's clinical role – consultant, ward nurse, manager, etc – needs to be represented separately from their qualifications – MD, RN, BHCS, etc – because even though roles usually follow qualifications, there is cross-over and exceptions.

GALEN formulates this principle as the 'principle of orthogonal taxonomies' which is extensively discussed in [24, 42]. In summary, there are two criteria:

a) Every elementary concept must have one, and only one, elementary parent. All multiple classification must be inferred from descriptions and definitions.

b) The elementary children of an elementary concept are considered disjoint but non-exhaustive – ***e.g*** if the children of 'Bone' are a list of specific bones and categories of bones, the bones and subcategories of bones must not overlap, but it cannot be assumed that the list is complete.

Our experience is that the principle of orthogonal taxonomies is key to re-use. For any specific use, ontologies may be developed using other principles, ***e.g.*** requiring that the entire taxonomy be a pure hierarchy, ***e.g.*** [9, 11, 13, 43]. However, to do so means making forced choices which are not represented explicitly. Anytime choices are made and the basis of those choices left implicit, the choices will be appropriate for one application but not another.

2.1.5 Normalisation and canonical forms

Normalisation is the process of reducing an expression to a standard, usually simplest, form. There are two levels of normalisation which must be dealt with:

a) Logical – ***e.g.*** to reduce the "fracture of a Long bone located in the femur" to "Fracture of femur". This is a purely logical operation dependent on the representation of 'Fracture', 'Bone', 'Long bone', and 'femur'.

b) Pragmatic – ***e.g.*** to agree whether a particular procedure is to be represented as the equivalent of "Fixation of Femur by means of insertion of pins" or "insertions of pins to fixate femur" [44]. In this case there is no simple logical solution and the problem must be solved by conventions and guidelines which are outside the formal model.

2.1.6 Separation of knowledge, meta knowledge, and representation knowledge

There at least five kinds of knowledge to expressed concerning any category in a concept representation:

a) *Domain knowledge* which defines and describes the category in terms of properties shared by all of its instances – *i.e.* as how we conceptualise the real world. .
b) *Meta knowledge* about the category itself – its use in language, whether or not it should be presented in a given situation, whether it is difficult to understand, common or rare in particular contexts, understood by particular users, etc.
c) *Representation knowledge* about how the category is represented in a given model, *e.g.* its immediate parent's and children in a given, model, whether it is elementary or defined, etc.
d) *Editorial knowledge* about the production and maintenance of the representation of the category – its source and authority, who entered and edited it, etc.
e) *Implementation knowledge* about how the representation is actually implemented in software.

The standard semantics for description logics is defined only for a) domain knowledge. GALEN's Common Reference Model itself concerns only domain knowledge, although there are 'kluges' for handling limited meta-knowledge and the tools provide means of dealing with representation knowledge and editorial knowledge.

2.2 Technical issues for design of an ontology

2.2.1 First class entities and modifiers

The concepts in ontologies tend to divide into two kinds:

a) Those which represent things which can exist on their own, *e.g.* physical objects, processes, ideas, etc.
b) Those which only make sense when linked to some other object as a modifier, modalities, or collections. Modifiers are notions such as "severe", or "soft", or "short" which describe first class concepts and specialise them further. Modalities are notions such as presence, uncertainty, family history etc. which take their meaning from the kernel first class concept. Sowa [45] after Pierce terms the first class objects simply 'firsts' and the modifiers 'seconds'. (The nominalised relations between the two – see below – he terms 'thirds'.)

The most important practical difference in a clinical ontology is that:

a) Lists of first class entities are almost always 'open', *i.e.* they cannot be assumed to be complete so that it is not legitimate to infer from the negation of some that one of the others is present. Lists of diseases, operations, and even bones and joints, are almost never complete. Reality is fractal and more detail can almost always be added.
b) Lists of modifiers may be 'closed', *i.e.* may be assumed to be complete so that inference of the form "not raised or normal, therefore depressed" can be justified logically, though they must be used with care clinically.

GALEN treats all concepts as 'open'. It never makes inferences such as "not absent implies present" on the grounds that they risk imputing a degree of logical rigour to clinician's statements which is rarely intended let alone achieved.

2.2.2 Nominalised relations or 'Features'

The choice of what should be an semantic link and what should be a concept is less simple than it seems. In any definition of the form

Disease <u>which</u> hasSeverity severe

the link can be nominalised to give an expression of the form:

Disease <u>which</u> hasFeature(Severity <u>which</u> hasState severe)

This is a general feature of any representation of this type, be it description logic, semantic networks, or conceptual graphs. Given that this transformation is always possible, at one extreme a system could be built with just two semantic links – hasFeature and hasState – so how should the decision be made? There are two criteria:

a) Need for further description. In most formalisms, semantic links can be organised into a kind-of hierarchy, but they cannot themselves be described. Therefore, if a the 'fact of being linked' may need to be described, even if in only some cases, then the link must be nominalised to a feature.
b) Consistency of representation. If there are a series of properties which appear analogous, it is almost impossible to maintain a system in which some are represented as links and some as features.

2.2.3 Dualities and qua-induction

If we define “mother” as the “female parent of a child”, we imply the possibility of describing a “child with a female parent”. (This process is sometimes called ‘qua induction’.) Many medical concepts come naturally in such dualities, and it is not always obvious which should be represented as primary. For example the ‘process of ulceration’ produces ‘ulcer lesions’. Should the process be defined in terms of the lesion or visa versa? or should both be treated as elementary and related? or should both be treated as aspects of some broader abstraction which we might phrase ‘Ulceration Situation’? The choice is unclear, but it needs to be made consistently.

GALEN has chosen to have just two concepts in all such cases corresponding to the process and resultant lesion, to make the process elementary, and to define the lesion in terms of the process. Having just two concepts has the disadvantage that a choice must always be made as to whether reference is being made to the process or the lesion. This puts a heavy load on the language and intermediate representation modules to make the correct interpretations.

2.2.4 Top level ontologies

The original belief of those developing the GALEN ontology was that it would be built from the bottom up. The top level ontology was seen as making little difference to classification and inference. Experience has largely confirmed this view technically but, paradoxically, refuted it pragmatically in terms of the human factors of development process. Regardless of its effect on inference, an agreed top level schema has proved essential to allow groups to co-operate to build and maintain a large ontology consistently. People need to know where to put things and where to find them, even if there are no inferences which depend on those choices.

However, just as all ontologies are approximations, all high level ontologies are to some degree arbitrary. Our conceptualisation of the world does not break down into a sequence of disjoint partitions. There are several existing starting points – PENMAN [46], Cyc [17, 47], traditional schemes deriving from Shank [48] and others in the AI/Linguistics field, and those deriving from Pierce via Sowa [45]. GALEN’s is adapted from that of Lenat and Guha.

In addition, it is our experience that each major field – clinical medicine, medical care management, protein/DNA function, art-history – which we have tackled with these tools requires one or two very high level abstractions which are very broad disjunctions which cut across the traditional boundaries of top level ontology. In GALEN, the concept representation *Phenomenon* and attribute *involves* are designed to cover anything which might be considered a disease, disorder, or condition of an organ or physiological process.

2.3 Ontological issues which interact with the semantics of the formalism

2.3.1 Negation and uncertainty

Negation and uncertainty pose difficulties for at least four reasons:

a) The meaning of negation and uncertainty in clinical medicine is unclear. For example, if there is no mention of diabetes in the record at all, what should be the answer to a query “Does the patient have diabetes?” Most database systems would answer “no” on the basis of a ‘closed world assumption’ and ‘negation as failure’ – the assumption that all relevant information about the domain of discourse is contained in the database and that therefore failure to find a fact can be taken as equivalent to its negation. In many clinical applications, neither assumption seems safe. Furthermore, if uncertainty is catered for, should it be included with negation or be a separate dimension? ***e.g.*** what are the comparative meanings of “possibly present” and “possibly absent”? .

b) The scope of the negation is often unclear. Three cases must be distinguished: a) “The patient does not have X”; b) “The patient has non-X”, and c) “The patient has X but not some specific kind of X”. The common cases in medical records are a) and c).

c) Adding negation and uncertainty to formalisms increases their computational complexity and makes normalisation to a unique canonical form difficult, and sometimes impossible. (Note that even if the formalism supports negation, the designers of the ontology may decide not to use it. KRSS which underlies SNOMED-RT supports negation, but it is the authors’ understanding that negation is not used in the SNOMED ontology.)

d) Negation and uncertainty are often represented in the information systems which hold the concept representation from the ontology. If negation can be represented both in the information system and the ontology, then the meaning of all possible combinations of negation in the two systems must be defined?

GALEN uses a formalism without negation and includes constructs such as ‘presence’ and ‘absence’ which can be qualified by a certainty (see *ClinicalSituation* below). It uses a separate mechanism for distinguishing normal/non-normal and physiological/pathological to cope with significant negatives such as ‘absent pulses’.

2.3.2 Defaults, indexing, and doubly universal definitions

The definition for 'A is a kind of B' in all formal representation is that 'All As are Bs'. Classifiers work by proving from the definitions, that anything which is an 'A' must be a 'B' (or equivalently, that any attempt to construct a definition for something which is an 'A' but not a 'B' will encounter a contradiction). Hence all of the properties in the definition and descriptive axioms about 'Bs' must apply to all 'As' without exception. Otherwise there is some 'A' which either i) does not meet the definition of 'B', or ii) contradicts some axiom about all 'Bs'. Unfortunately, adapting such inferences to accomodate default reasoning has so met with at best limited success, both logically and computationally [49].

Therefore, in GALEN or any other system based on description logic, statements in definitions and descriptions are 'indefeasible'– *i.e.* must apply to all subconcepts without exception. This contrasts with the frame systems from which such representations were derived, in which default values play an important role.

However, a description-logic-based ontology can still help with reasoning about defaults by gathering together the set of all the most specific candidate values. (One candidate value is more specific than another if the concept from which the more specific is inherited is a kind of the concept from which the less specific value is inherited [50].)

GALEN's experience is that if the taxonomies are properly orthogonal, the set of candidate values will usually have exactly one member. If it has more than one member, then GALEN treats this as a signal that other reasoning methods and knowledge are required. Depending on the application these additional methods may range from simple union to more complex systems of priorities to entire rule based systems or belief networks.

Another way of looking at this technique is to treat the ontology as an index to the default information. One of the major advantages of ontologies using classifiers is that they can provide indexes of any required granularity and cope with as many special cases as required. For example in a drug ontology, a single scheme can index side effects whether they refer to broad classes such as beta blockers or specific brands of specific drugs.

Finally, many 'default' statements are of the form 'all Xs haveLinkTo all Ys' or rather 'generally Xs haveLinkTo Ys', *e.g.* "Generally beta blockers are contraindicated in Respiratory disease". Standard description logic based classification algorithms do not support constructs of the form 'all Xs haveLink all Ys' directly, even leaving aside the complication of defaults. Treating defaults as facts to be indexed has the added benefit that it plugs this gap.

GALEN refers to such indexed statements as 'extrinsic' because they do not affect classification within the ontology proper.

2.3.3 Normative statements, congenital malformations, and imputed intentions

Many of the descriptive axioms used in terminology models are actually 'normative' rather than absolute, *i.e.* they really pertain to our view of 'normal' anatomy, physiology etc. This gives problems when describing congenital malformations and mutilations. There are three complementary approaches to this problem:

To adjust the interpretation of the semantic links and concepts. For example, GALEN interprets the *hasDivision* link in such a way that 'Hand isDivisionOf Arm' is true, even if the hand is severed from the arm. Since we can still talk about the missing hand, this is the best 'logical approximation'.

To relax the constraints such as "every hand has only five fingers", or "the left lung has no middle lobe", but flag the presence of the normal parts as 'normal' and the other parts as abnormal. GALEN uses this approach extensively – see *normal/nonNormal* below.

To model both normal and abnormal but use the interface and 'perspectives' to limit the normal view only to the normal conformation.

Normative statements give more difficulty when applied to procedures or other areas outside anatomy and physiology and can be closely tied to issues of pragmatic normalisation. To take the classic example used by the O'Neil [44] "Insertion of pins in the Femur". This operation is almost certainly performed in order to fixate the femur. To classify this operation under "Operations to fixate long bone" this information must be added to the description of this category of operations. To do so risks imputing unstated intentions to the clinicians? For example, appendicectomy is usually performed to cure appendicitis but may be performed prophylactically, *e.g.* before a polar expedition. GALEN is extremely cautious about making normative statements in the Common Reference Model, preferring to manage them through the external perspectives, intermediate representation and language modules. However, in developing classifications of surgical procedures, some normative statements seem to be required to meet clinicians' expectations.

2.3.4 The attribute hierarchy

In GRAIL, as in many but not all description logics, the one attribute (role) can be a kind of another just as one category can be a kind of another. There are several key uses for the attribute hierarchy, but three deserve special mention and are supported by naming conventions in GALEN:

a) To allow single-valued and multi-valued variants of the an attribute. Logically, the single-valued attribute must always be a descendent of the multi-valued attribute, and its presence is signalled by the naming convention that the single-valued variant contains 'specific' or 'specifically', e.g. hasSpecificConsequence, or actsSpecificallyOn. These 'specific' attributes are often used to indicate a main, or primary action, cause, etc. – see "multiple causation".

b) To allow direct and transitive variants of an attribute, ***e.g.*** *hasComponent* and *hasDirectComponent* and *hasComponent.* In this case facts almost always involve the 'direct' variant and queries the transitive variant.

c) To allow very general queries, such as "disorders of heart". The *involves* attribute (see "constructive attributes" below is the best example of this.

2.3.5 Transitive relations,

Parts and wholes, causal links, and connections are all transitive and some characteristics are 'inherited' along these transitive relations. Establishing the pattern of transitive relations and inheritance along them is a key part of any ontology of medicine – see *Part-whole relations* below. (GRAIL's formal mechanisms related to transitive relations are described in [1].

2.3.6 Concept constancy in compositions and avoiding 'Up-to-ism'

If a concept's representation depends on its use, then this limits re-use. Concepts such as "lobe of the liver" or "fluid in cyst in the kidney" should appear the same regardless of context – whether as aspects of disease, targets of surgery, substance to be injected or drained, or specimen in a pathology examination. Since many of these concepts are themselves composite, GALEN must allow definitions to be embedded within definitions to any degree required – ***i.e.***GALEN supports expressions such as "upper part ***of*** third segment ***of*** middle lobe ***of*** right lung" to any depth and their subparts appear the same in the composition as when isolated.

Arbitrary limits on the complexity of expressions also limit re-usability. Arbitrary limits are sometimes caricatured as 'up-to-ism' – you can have 'up to' some limit of a particular option. Traditional enumerated medical terminologies have suffered seriously from 'up-to-ism' with limits on the depth of the hierarchy and the number of items at any given level. The most common source of 'up-to-ism' in logical knowledge representations in medicine arises from the handling of locations and sites, ***e.g.*** Location-right lung, sublocation-middle lobe, sub-sub location-third segment. GRAIL's support for transitive relations means that definitions avoids the need for such constructions.

2.3.7 Not 'lying to the system'

All logical representations are approximations, but logic guarantees only that true conclusions follow from true premises. If restrictions in the formalism cause users to 'work around' it to the extent that some definitions and descriptions are false, then even if the result serves well in one application its general behaviour is difficult to predict. The most common example of 'lying to the system' is the use of 'is kind of' instead of 'is part of'' which works, for the most part, as long as the application involves only disorders and procedures, but breaks down for more general descriptions of pathophysiology (see [23]).

2.3.8 Adjacency, wholeness, and spatial (and temporal) reasoning

"*A* fracture of the tibia and fibula" makes sense; "*A* fracture of the tibia and humerus" does not. Furthermore, "A fracture of the tibia" is not a kind of "Fracture of the tibia and fibula" – which it would be by GALEN's rules if we simply considered "tibia and fibula" as a single structure with parts "tibia" and "fibula". In an anlogous case, "inflammation of the large bowel as a whole" is different from "An inflammation of the large bowel"; furthermore, "an inflammation of the ascending colon" is a kind of "inflammation of the large bowel" but not of an "inflammation of the large bowel as a whole". Dealing completely with the issues illustrated by examples requires a degree of spatial reasoning which is currently beyond what can be done easily in GALEN – or any other description logic – although there are several experimental 'work arounds'.

2.3.9 Numerical conversions, calculations, and other 'non-terminological' reasoning

There are numerous services which a user might naturally expect to be packaged with a terminology which require entirely different reasoning from logical classification based on definitions and descriptions. The most obvious of these is conversion between different unit and co-ordinate systems. Other examples include other forms of abstraction, ***e.g.*** from numerical values and normal ranges to qualitative properties such as elevated, normal, or depressed.

3. GALEN's Clinical Ontological Schema

The following sections relate major features in the GALEN ontology to the above considerations. A sketch of the ontology can be found in Appendix A, and the entire ontology is available at www.OpenGALEN.com.

3.1 Top Level Ontology

The remainder of this paper uses GALEN's nomenclature: concept representations are known as 'categories' and semantic links as 'attributes. By convention, the top level domain, in GALEN's case medical, category in any GRAIL model is *DomainCategory*.and the top level attribute is *DomainAttribute*.

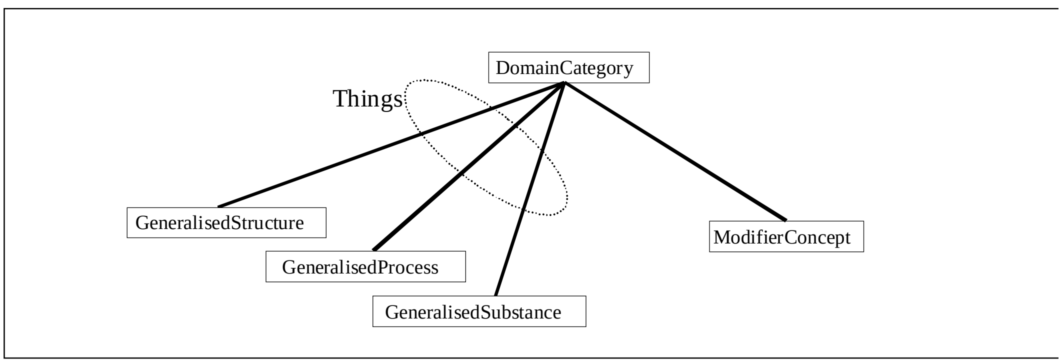
GALEN's top level distinction under *DomainCategory* is between first class entities, or *Things*, and everything else, or *ModifierConcepts*. *Things* are divided into

- *GeneralisedStructures* — abstract or physical things with parts independent of time
- *GeneralisedSubstances* — continuous abstract or physical things independent of time
- *GeneralisedProcesses* — changes which occur over time

This structure is adapted from Lenat and Guha. However Lenat and Guha maintain a distinction for processees analogous to that between 'Structure' and 'Substance', *e.g.* between "The digestion of a meal" and "The activity of digestion". In GALEN's experience, both users and knowledge engineers have found this distinction confusing. Too many medical processes have ill-defined beginnings and ends for it to be easy to maintain the distinction consistently. Hence no such distinction is made.

Note that because having a category labelled 'Thing' led to arguments, it was deleted so that in the GALEN Common Reference Model as published *DomainCategory* breaks down directly as shown in Figure 1.

Figure 1: Top Level GALEN Category Ontology

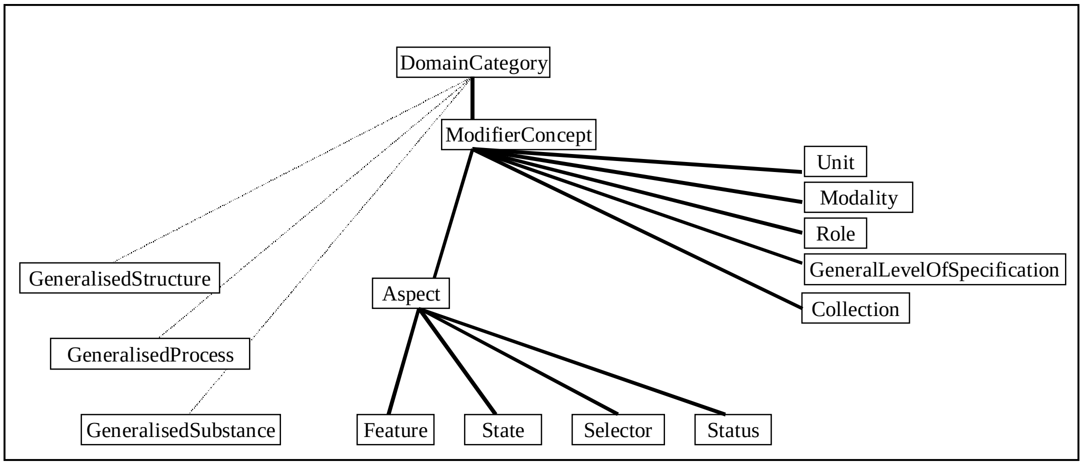


3.1.1 Modifiers

The first level break down of *ModifierConcept* is is shown in Figure 2 and falls into:

- Aspect and Modality
  - *Aspect* –' modifiers' proper', normally used in the pattern: *PrimaryThing which …modifier*, as described above in *First Class Entities and Modifiers* above
  - *Modality* – modalities as described under *First class entities and modifiers* above, normally used in the pattern *Modality which ….*
- Other concepts which are dependent on first class concepts for their meaning
  - *Role* – arbitrary concepts used to make elementary taxonomies orthogonal, *e.g. DoctorRole, HormoneRole, DrugRole* etc. See roles and role designating attributes, below.
  - *Collection* – set, system, etc. GALEN has a weak notion of *Collection*. There are no special features in the formalism to support operations on collections.
- Other things that have special significance or behaviour
  - *Unit – mg, day, etc.*
  - *GeneralLevelOfSpecification – a category of meta-knowledge which is 'kuged' into the system to help guide certain applications but which is now largely disused.*

Figure 2 : Modifiers in High Level GALEN Category Ontology



Aspect is further broken down into:

- *Features* – nominalised relations between such as 'Level' or 'Severity' which must be modified by one or more States to have a meaning in a 'Feature-State' pair.
- *States* – the 'values' of *Features, e.g.* 'mild', 'moderate', or 'severe'
- *Selectors* – .values of selections, largely in anatomy, such as 'left/right' or 'upper/middle/lower' as opposed to 'to the left of' and 'to the right of', etc. . Selectors identify a specific entity rather than modifying it.
- *Statuses* – modifiers which are used in the 'internal' workings of the model such as *'normal/nonNormal'*, *'countable/indefinitelyDivisible/mass'* and various topological indicators. (There are also a few things which should be 'states' but have never been moved because it makes little practical difference'.)

All mutable properties are expressed by *feature-state* pairs, as described in *Nominalised Relations* above, *e.g.*

Disease which hasFeature (Severity which hasState severe)

By contrast, *Selectors*, and *Statuses* are always linked to the entity they modify by a single semantic link, *e.g.*

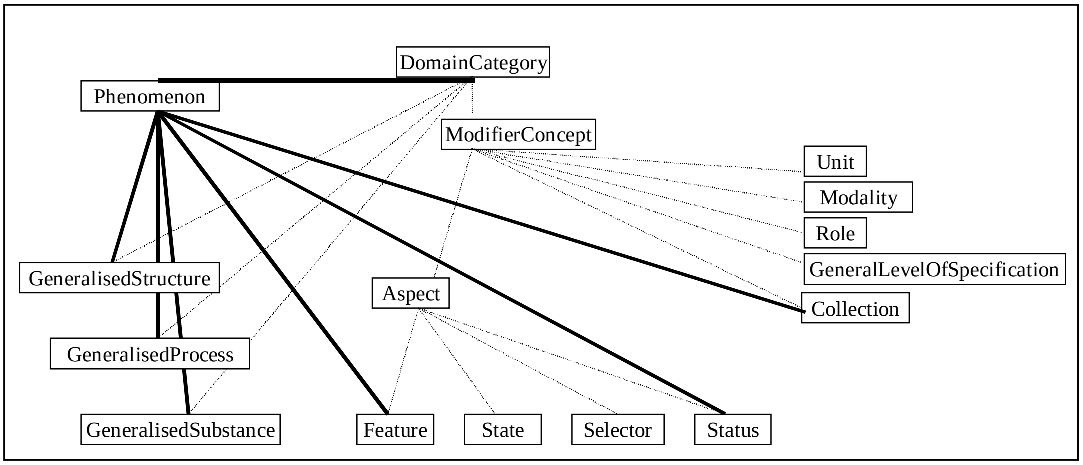
Hand which hasLeftRightSelector rightSelection

There can never be anything more to say about the 'rightness' of the right hand, it just is; likewise status such as nonNormal which in GALEN act simply as 'flags' – see *Normal/NonNormal…* below.

3.1.2 Phenomenon – Secondary High Level Ontology

As noted in *Top Level Schema* above, most ontologies seem to require at least one very high level disjunctive category to represent key concepts. In GALEN this is *Phenomenon*. A *Phenomenon* is anything which can be, or be modified to be, worth noting clinically as nonNormal or pathological. As shown in Figure 3, *Phenomenon* is the disjunction of *Thing (GeneralisedStructre, GeneralisedProcess, GeneralisedSubstance), Feature, Status,* and *Collection*. This is really too inclusive for a general ontology, since it allows things like rocks to be 'pathological', but the enormous effort to tighten the constraints has not so far been warranted in a clinical ontology for clinical applications.

Figure 3 : Secondary High Level Taxonomy



3.2 High level attribute ontology

3.2.1 Top Level attribute ontology

The top level of *DomainAttribute* breaks down into:

ConstructiveAttribute -	Links between first class entities, *i.e. Things*, *i.e.GeneralisedStructures*, *GeneralisedSubstances*, and *GeneralisedSubstances*
ModifierAttribute -	Links between *Things* and *ModifierConcepts*
TemporalAttribute -	Links between processes and statuses via temporal relations (not fully complete)

3.2.2 Constructive Attributes

The key constructions in most medical terms are to locate a disease or procedure in a structure or part of a structure. The most common schema in GALEN for disease or procedure concepts, and by far the most common schema for queries or abstractions is:

Disease/Procedure <u>*which*</u> *LocativeAttribute BodyStructureOrProcess*

LocativeAttribute has been steadily generalised in the course of the project until, like Phenomenon, it is effectively a domain specific disjunction of a number of other key attributes. Because of its generality, it has the alias *'involves'*.

ConstructiveAttributes therefore break down into three main attributes plus the domain specific disjunction, *LocativeAttribute* *(*alias *involves)*:

PartitiveAttribute -	Part-whole relations as discussed further under anatomy and processes.
StructuralAttribute -	Non-partitive structural relations such as connects, passes through, contains, etc.
FunctionalAttribute -	Attributes involving functional relations.
'involves' (LocativeAttribute) -	Links lesions, processes, and procedures to their location. Tumours are 'located in structures' rather than being 'part of structures' in the GALEN ontology. Also subsumes Process *LocativeAttribute* and *isCausallyLinkedTo,* and *isFeatureOf*

It may seem odd to think of causal attributes as 'locative', but, for example, finding 'spider haemangiomata' under 'Phenomenon involving liver disorder' is what would be expected.

3.2.3 Modifier Attributes

The modifier attributes and modifier categories are intimately tied, one main branch of the attribute hierarchy for each branch of the modifier hierarchy: *hasFeature, hasState, hasStatus, hasSelector, hasModality*

It is an artefact of the GRAIL language that each of these attributes has a multitude of child attributes, one for each feature-state pair, status set, selector set, etc. because whether an attribute is single-valued or multi-valued is a property of the attribute rather than the attribute-value pair. Hence in order to allow a disease to have only one SeverityFeature but still to have many different other features, a separate single-valued attribute is required for each feature. For example, the raw internal form of the reference model for a “severe disease” is:

Disease which hasSeverityFeature (Severity which hasSeverityState severe)

3.2.4 *Roles* and *RoleDesignatingAttributes*

The principle of orthogonal taxonomies leads to a wide range of specialised hierarchies which are used to designate aspects of concepts which themselves are best seen as natural kinds. In GALEN these are all known as ‘roles’ and linked by attributes named playsXRole or hasXRole. The range of usage extended from marking substances as hormones using *homoneRole* to identifying professions by playsClinicalRole. A full list is included in Appendix A.

3.2.5 Breaking up long lists and the *NAMED…* convention

The principle of orthogonal taxonomies combined with a principled approach to choosing natural kinds, results in a broad flat hierarchy of elementary categories which is difficult to work with. For convenience GALEN breaks this hierarchy up into units which are convenient for the developers by introducing categories such as *NAMEDBodyPart*, *NAMEDDrug*, etc.

3.2.6 Transitivity and *specialisedBy* axioms

A key part of the high level ontology is the decision as to which attributes are transitive and which other attributes can be inherited along other attributes – indicated by the GRAIL ***specialisedBy*** axioms. The easiest way to understand the effects of these axioms is as principled statements such as “Components of divisions are components of the whole, but divisions of components are not divisions of the whole”, and such statetements are used throughout the following sections. A summary of the main specialisedBy axioms themselves can be found in Appendix A.

3.3 Anatomy

3.3.1 Part-whole relations and physical connection

There has been much study of parts and wholes, or ‘partitive relations’, and an entire field of ‘merology’ is based on their study. Two useful general references are [51, 52]. Two specialised references vis a vis discription logics are [53-55]

GALEN’s principles are:

a) All partitive relations are transitive.
b) Diseases/disorders/procedures of/on a part are pertain also to the whole.
c) Partitive relations break down into two groups: those involving parts of physical objects and those involving constituents in mixtures of substances or granular material. (Collections are not considered ‘partitive’ in GALEN though they are in [51, 52])
d) Partitive relations involving physical parts occur in three main ways:
   - Solid and Surface divisions – things which are roughly self similar, at least to the extent that they have similar layers, ***e.g.*** the upper and lower arm are divisions of the upper extremity. The distinction between solid and surface divisions is roughly parallel to the distinction between two- and three-dimensional reasons in he Digital Anatomist Project [56].
   - Layers – things like skin or the muscle or periosteum which occur in all divisions of an object. GALENs approximation is that a layer of a division is a kind of a layer of the whole. The more correct relation as pointed out in [53] is that a layer of a division should be a division of a layer of the whole – ***i.e.*** the skin of the hand should be a division of the skin of the upper extremity.
   - Components – things like joints, ligaments, processes, organs, which occur only in one (or occasionally more) divisions of an object. Layers of components are not layers of the whole – ***i.e.*** the surface of the cusp of a heart valve is neither a kind of nor a part of the surface of the heart.
e) Connection is transitive but not partitive; ‘branch-of’ is neither partitive nor transitive (otherwise all arteries would be branches of the aorta – perhaps true in some sense but not very useful)
f) Connected physical sets such as the ‘digestive tract’ are distinct from functional systems such as ‘the digestive system’. For example, the pancreas is part of the digestive system but not of the digestive tract.

The general principles for transitivity for partitive attributes are:

a) All partitive relations are transitive
b) Components of any kind of parts are components of the whole
c) Layers and divisions of divisions are layers of the whole
d) Layers and divisions of components are *not* layers or divisions of the whole

The attribute hierarchy which achieves this is given in Appendix A.

3.3.2 Topology, cavities and containment

All structures in the GALEN common reference model have a topology which may be ***hollow*** or *topologicallySolid.* Being solid is simple. The key feature of being hollow is that any hollow object *definesSpace* known as a *Cavity*. Things can only be 'contained' in the cavities formed by hollow objects.

However, being ***hollow*** is actually quite complex and breaks down into

surfaceHollow –	Surface regions such as the 'abdomen' which overlie a cavity and are often said to have things in them. This may be a 'kluge' but it is difficult to cope with many common medical expressions without some such construct.
trulyHollow –	Properly hollow structures.
closedHollow –	No openings
tubularHollow –	One or two openings. The cavity is a *lumen*.
bilayered –	Membranes which form potential spaces such as the peritoneum or pleura

Containment is not considered partitive in the GALEN common reference model because it is not consistently the case that 'disorders of a contained thing are a disorder of the containing thing' – "disorders of the heart" are not usually considered "disorders of the chest", although "surgery on the heart" is considered "chest surgery" but "catheterisation of the heart" is not considered as "catheterisation of the chest". (To capture "heart surgery" as "chest surgery" requires a normative statement that "open surgery on the heart" requires "opening the chest".) There is a family of 'partitive containment' attributes to cover the mixed case.

3.3.3 Tissues, Cells, and substances: *mass*, *discrete*, and *indefinitelyDivisible*

Most western languages make a distinction between 'mass nouns' such as which are normally used in the singular "water", "sand", and "shopping" and 'count nouns' such as "stick", "stone", and "grocery" which allow plurals (and occasionally take the singular only rarely). The example of "the shopping" and "the groceries" shows that this is a linguistic rather than a semantic distinction, but there is a corresponding semantic distinction ***e.g.*** between 'stuff' and 'things' in Lenat and Guha [17].

In the GALEN Common Reference Model, structures and substances have a *Countability* which can be one of:

discrete –	*individual bones, organs, membranes, etc.*
mass –	*substances and tissues*
indefinitelyDivisible–	*Cells, grains of sand, etc.*

The *indefinitelyDivisible* category was added to cover things like cells which are usually treated en masse as in their count-concentration in a body fluid, but which can have discrete parts. It can be argued that it would be better to use the *Multiple* variant of collection described under diseases and use a construct equivalent to "the count concentration of a multiple of red blood cells in blood" rather than "the count concentration of red blood cells in blood", but at the time the decision was made, there was insufficient evidence of either the advantages or the consequences to justify the change. Relatively little use is made of these constructs in practice, so the question remains open.

The partitive attribute for mass and indefinitelyDivisible attributes are *isMadeOf*. The containment (non-transitive) attribute is *isMixedThroughout* .

3.3.4 Selectors

Selectors are discussed under 'Modifier Attributes' above. A more extensive list of selectors is given in Appendix A.

3.3.5 Regions

The problem of describing the regions of the body is one of the significant headaches for any system attempting to describe anatomy logically.

a) Regions have ill defined borders

b) Regions can be either two-dimensional surface regions or three-dimensional solid regions, and the distinction is not always clear.

c) Regions are often named identically with the primary structure which they contain: the region of the lower extremity associated with the "knee joint" is the "knee region", and furthermore the surface region indicated by the knee is often referred to interchangeably with the three dimensional region.

d) The thing for which a region is named is clinically significant but not easily represented in the ontology – there are many structures in the left anterior chest besides the heart, but the "precordium" is specifically associated with the heart.

e) The definitions of some regions, such as the axilla and perineum, varies amongst authorities.

The final problem is obviously beyond the scope of GALEN, in one sense, but presents problems when attempting to compare GALEN's anatomy model with others – a discrepancy can be due either to errors in modelling or differences in the authorities used. Errors in modelling can be easily fixed; differences between authorities must either be reconciled or separately represented. A significant number of the discrepancies found in comparing the GALEN model of surgical procedures and that of the UK Clinical Terms version 3.1 were found to originate in differences between authorities [57].

GALEN represents two- and three-dimensional divisions into regions using two families of attributes – *hasSolidDivision* and *hasSurfaceDivision*. (*i.e.* solid/3-dimensional regions) and correspondingly *SolidRegions* and *SurfaceRegions.*

Regions are typically named for a single specific body part, and the "periaortic space" is described by *hasSpecificProximity Aorta.* Similarly, the contents of ill defined spaces and nodes are described analogously to the *AxillaryLymphNodes* by *hasSpecificProximity Axilla.*

3.3.6 Bits and Pieces

Terms such as 'capsule', 'spine' or 'edge' are widely used in anatomy to identify anatomical substructure elements –*e.g.* "capsule of kidney", "spine of 5th lumbar vertebra", "edge of liver" etc. Each such term can be ascribed at least some level of definition although such definition may be less than precise.

In modelling such anatomical substructure there are two choices.

a) To represent the generic notions as natural kinds and the real anatomic structures as defined composites, *e.g. Angle <u>which</u> isStructuralComponentOf Mandible, Pole <u>which</u> isStructuralCompoenentOf Kidney, etc.*

b) To represent the substructures can themselves be taken as natural kinds, *e.g. PoleOfKidney, PoleOfOvary*, etc. with no explicit relationship to the more abstract notion of e.g. *Pole*

In general, GALEN has chosen the compositional method because there seems to be sufficient commonality in notions such as "lobe", "pole", "segment", etc. to merit capturing them individually.

3.3.7 Other anatomical notions represented

In addition to whether objects are solid or hollow, mass or discrete, GALEN three other anatomical notions:

a) SurfaceVisibility – whether a structure is internal or external

b) Shape – laminar, linear, etc.

c) Paired or unpaired – which includes a flag as to whether the pairing is 'mirror imaged.

Appendix A includes the full list of such attributes and values.

3.4 Diseases

3.4.1 Normal/NonNormal and Physiological/Pathological

What is a "disease" or "disorder"? What does it mean to say that something is "normal" or "abnormal"? "pathological" or "physiological"? These issues generated great debate early in the course of the GALEN programme. Given many different philosophical definitions, the only recourse was to identify what operational outcomes were required to provide a reasonable logical approximation. These include:

a) Distinguishing normal anatomy from abnormal and to list the normal anatomical parts, connections, etc. of any structure.

b) Identifying those things whose presence was potentially noteworthy in a medical record or similar.

c) Flagging things as clearly 'diseases' or 'pathological', *i.e.* something close to "in potential need of medial management"

d) Representing the notion of being 'abnormal but not pathological' which we took as meaning something like "note-worthy but not in need of medical management".

e) Recognising that the presence of some things is always pathological, *e.g.* a malignant tumour or a fracture

To achieve these objectives, GALEN defines two independent status distinctions, *normal/nonNormal* and *pathological/physiological* plus two rules enforced through GRAIL's necessary statement mechanism:

a) anything *pathological* is *nonNormal*, and

b) anything *normal* is *physiological.*

In addition it defines a further set of statuses: *intrinsicallyNormal/intrinsicallyNonNormal* and *intrisicallyPhysiological/intrinsicallyPathological* plus the rules:

a) Anything *intrinsicallyNonNormal* is *nonNormal*

b) Anything *intrinsicallyPathological* is *pathological*

(The converses are not true, things which are *intrinsicallyNormal* are not necessarily *normal* – they may have acquired disorders or abnormalities; likewise for *intrisicallyPhysiological* and *physiological*)

Using this scheme, normal anatomical structures are those which are *intrinsicallyNormal;* noteworthy things those which are *nonNormal*, diseases and disorders things which are *pathological*, and things like tumours which are always diseases *intrinsicallyPathological.* The combination *nonNormal* & *phyiological* is allowed but *normal & pathological* disallowed.

The label 'nonNormal' was arrived at after much discussion as giving rise to fewest preconceptions of the labels suggested. There remain in the GALEN model a number of specialisations of *nonNormal* including *unusual*, and *variant*, which were left in to satisfy those who felt that further distinctions would be necessary. However, in practice these have rarely been used and their practical value is questioned.

Figure 4: Alternative formulation of Abnormality and Pathological Status

AbnormalityStatus
  nonNormal
    pathological
      intrinsicallyPathological
    intrinsicallyNonNormal
      intrinsicallyPathological
IntrinsicStatus
  intrinsicallyPhysiological
    intrinsicallyNormal
  intrinsicallyNonNormal
    intrinsicallyPathological
	nonNormal = ¬ *normal*
	pathological = ¬ *physiological*
	intrinsicallyNonNormal *disjointFrom* *intrinsicallyNormal*
	intrinsicallyPathological *disjointFrom* *intrinsicallyPhysiological*

physiological *normal*	

Enforcing 'rules' using GRAIL's necessary statement mechanism is inherently expensive computationally. Modern description logic classifiers based on tableaux algorithms [58, 59] with negation anddisjunction can infer from *A subsumes B* that ¬*B subsumes* ¬*A*. Using systems, essentially the same result can be obtained using the hierarchies in the top half of Figure 4 from which the hierarchy in the lower half can be inferred. (The 'intrinsic' forms are declared disjoint but not negations, because otherwise it would be inferred that *intrinsicallyNormal* subsumes *normal* and *intrinsicallyPhysiological* subsumes *physiological* – an inference with which limits the usefulness of *intrinsicallyNormal* as a 'flag' for normal anatomic structures and physiological processes.)

3.4.2 *Phenomenon* and "Disease"

What then is a "disease"? The closest logical approximation to "Disease" or "Disorder" in the GALEN Common Reference Model is *PathologicalPhenomenon* defined as:

Phenomenon which hasPathologicalStatus pathological

The closest logical equivalent to "Disease of Organ or System" is:

PathologicalPhenomenon which involves OrganOrSystem

For example, "cardiovascular disease" is represented by the GALEN concept *CardiovascularPathology* which is defined as:

PathologicalPhenomenon which involves CardiovascularSystem

The GALEN category *Phenomenon* and attribute *involves* have been carefully crafted to try to capture the various ways in which things can go wrong with organs or processes to constitute "diseases of …" or "disorders of …". The label *PathologicalPhenomenon* explicitly avoids implying too close a mapping to any natural language phrase such as "disease", "disorder", or "condition".

3.4.3 Causation

Causation is a critical notion to medical concepts but surprisingly slippery. The GALEN common reference model recognises at roughly four dimensions.

a) Strength of association – from statistical to physiological cause
b) Immediate vs late
c) Whether thought of as a 'complication' or a 'cause'
d) Whether conceptualised as the primary or secondary cause as indicated by whether it is used in the naming or not.

These dimensions are captured, approximately, by the hierarchy of causal attributes given in Appendix A. The complexity of the hierarchy has grown over the length of the project. Its current ramification makes a strong case for 'reifying' the attribute *isCausallyLinkedTo* to the category *Causation* to create a concept which can then be described, specialised, and classified along each dimension independently according to the principle of 'orthogonal taxonomies'.

The key transitive rules are that the attributes which indicate close causal connections, ***e.g.*** *isSpecificImmediateConsequenceOf* are transitive whereas attributes indicating loose connections such as *isLateConsequenceOf* or *hasAssociation* are not.

3.4.4 Multiple causation

Many conditions are defined by their cause, ***e.g.*** "viral pneumonia", "bacterial meningitis", etc. What is to be done about situations in which there is more than one cause? Clinicians dislike the logical inference that "mixed pneumonia" is a kind of "bacterial pneumonia" and at the inability to distinguish between a "mixed pneumonia" and a "viral pneumonia complicated by bacterial infection". Nor would any clinician want to imply that "diabetic renal failure" could not have other contributing causes. On the other hand, the notion of an "infections involving bacteria" certainly ought to include such mixed conditions.

GALEN's solution to this problem is to use a special child attributes of ***isImmediateConsequenceOf***, ***isLateConsequenceOf*** ***i.e.*** ***isSpecificImmediateConsequenceOf***, ***isSpecificLateConsequenceOf***, etc.

Hence *ViralPneuomnia* can be defined as *Pneumonia* <u>*which*</u> *isSpecificImmediateConsequenceOf ViralInfection* while still be specialised by further criteria such as ***isImmediateConsequenceOf*** *BacterialInfection*. (By the GALEN naming conventions, the 'Immediate' refers to time and not importance.)

3.4.5 Non-localised diseases

There are a range of diseases which do not fit the patterns above – ***e.g.*** diabetes, the collagen vascular diseases, and perhaps even hypertension. These are systemic diseases which are not usually thought of as being localised to a particular part of the body or a particular physiological systems. Furthermore, they are broad syndromes of the consequences of some metabolic disorder, for example, should "diabetes" be identified with a notion such as "ineffective insulin mediated glucose metabolism" or with the syndrome involving welter of potential consequences ranging from retinopathy to leg ulcers which result?

GALEN's approach is to treat diseases such as "diabetes" as natural kinds (see above) of *IntrinsicallyPathologicalBodyProcess* and then to further describe them in terms of their causes and subprocesses. (By convention, the underlying mechanisms is currently usually described as a 'subprocess' rather than a 'cause', but this has little consequence for classification. What is important is that the systemic disease is distinct from the underlying pathology.)

3.4.6 Clinical Situations, presence and absence – encapsulating concepts for Medical Records

Two of GALEN's specific objectives as indicated in the introduction were to encapsulate concepts in so that they could be incorporated into traditional medical record architectures and to provide means of mapping to existing coding and classification schemes. What is it that must be entered in a record or mapped to a coding scheme? Answers to the two questions are similar but not identical.

For medical records which have been designed for traditional coding schemes:

a) It must be able to include the negative as well as positive concepts, since some medical records systems provide negation within their information model but others do not.
b) It is often a complex of several conditions – ***e.g.*** A with B without C
c) It is often necessary to record causal or temporal relations to other things in the medical record and co-ordinate that information with the concept – ***e.g.*** "nephropathy secondary to diabetes", or more difficult "fat embolism secondary to *this* fracture of the femur".

GALEN achieves a) and b) by 'wrapping' the kernel concept in two outer modalities:

a) *Existentiality – presence* or *absence*

b) *ClinicalSituation* – a clinical 'chunk' to be recorded and treated together.

Hence the total expression for a concept representation for recording in a medical record or mapping to a coding system is always of the form of the example below: :

ClinicalSituation <u>*which*</u> *isCharacterisedBy* <*presence* <u>*which*</u> *isExistenceOf StomachUlcer*
presence <u>*which*</u> *isExistenceOf StomachPenetration*
absence <u>*which*</u> *isExistenceOf Haemorrhage*>.

Key principle for transitivity so that this works is that a *ClinicalSituation* which is characterised by the *presence* or *absence* of some condition, is itself characterised by that condition – hence the above example would be subsumed by *ClinicalSituation* which involves *Stomach*. (Note that for the pathology, it is essential to specify whether the involvement is of the *presence* or the *absence* of the condition)

3.4.7 Levels

A recurrent pattern pointed out by Shahar [60] is in handling departures from the expected or normal value and changes in states. GALEN has adapted Shahar's scheme to provide a consistent ontology for all measurements which can be elevated, depressed etc. based on the re-usable feature *Level* whose potential values are illustrated in Figure 5.

Figure 5 Attributes and potential values related to the feature Level.

hasAbsoluteState	(red, cold, high)
hasQuantity	(an integer, and a unit e.g. 37 degreesCelcius)
hasChangeInState	(reddened, cooled, raised)
hasTrendInState	(reddening, cooling, rising)
hasRelativeLevelState	(redder, cooler, higher)
hasExpectedLevelState	(redder, cooler, higher)

3.4.8 Mapping to coding and classification systems

In general, this mechanism, and its analogue for procedures, suffices to produce concept representations which map directly to items in traditional coding and classification systems such as ICD9/10 or the disease axis of SNOMED-International with a few provisos.

a) The categories in the GALEN Common Reference Model do not represent codes directly, they are mapped to them using the indexing methods described in section 2 of this paper. The general rule is that a GALEN concept is mapped to the most specific code on which it is indexed. If there is more than one such code, then other mechanisms for disambiguation are required.

b) A code may be mapped to more than one GALEN category, especially if its definition has an 'includes' clause.

c) The 'most specific' mechanism copes with virtually[1] all of the 'excludes' cases – *e.g.* "hypertension excluding pregnancy" is dealt with automatically by the fact that elsewhere in the coding system there is code for the more specific category "hypertension with pregnancy".

d) Any code rubric including 'other' or 'not elsewhere classified' must be mapped to a artefactual term in GALEN which is specific to that coding system. In theory it would be possible to map other to a complex Clinical situation of the form: 'Presence of parent and absence of … list of all known children…' but this produces extremely large concepts which slow down the implementation. More seriously, the members of the project are unconvinced that these terms are used in such a consistently logical way.

e) Any code rubric including 'Not otherwise specified' ('NOS') is mapped to the parent concept with a suitable annotation on the mapping.

3.5 Procedures

The GALEN procedure schemas are covered extensively in [27]available on the *Open*GALEN web site at www.OpenGALEN.org. Two key issues are outlined below.

3.5.1 Direct and indirect objects: adaptation of the CEN env1828 model

GALEN took the CEN env1828 scheme for surgical procedures as its starting point, but the richer modelling environment and features of GRAIL allowed it to be significantly simplified. The original model is shown in figure 6. plus four constraints:

a) A surgical deed is seen as a part of an overall surgical procedure

[1] We know of no exceptions but are reluctant to make the statement categorical as yet.

b) A surgical procedure must have a direct object (pathology, anatomy or equipment)
c) A surgical procedure must have a surgical deed
d) A surgical procedure must have anatomy either as a direct or an indirect object

The direct and indirect object functioned much as location and sublocation discussed under 'up-to-ism' above. Because of GRAIL's ability to handle transitive relations, this could be simplified to the form shown in figure 6b, where anatomy could be either an entire organ, or an indefinitely embedded description, such as "upper segment of

This structure was further extended in three ways

a) A series of modifiers to cover open/closed, approach, the person performing the operation (significant to some remuneration schemes), means, and extent.
b) The possibility of subprocedures was added. This is a critical change as many, if not most, procedures in existing classification are defined in terms of a series of subprocedures.
c) An outer layer to accommodate *performance* or *nonperformance* analogous to *presence* or *absence* for diseases was added to account for the fact that some surgical procedure descriptions include "doing X without doing Y". The outer *SurgicalDeed* now acts analogously to *ClinicalSituation* – see *Diseases* above.

The resulting model is shown in 6c.

Figure 6a: CEN Model for Surgical Deeds

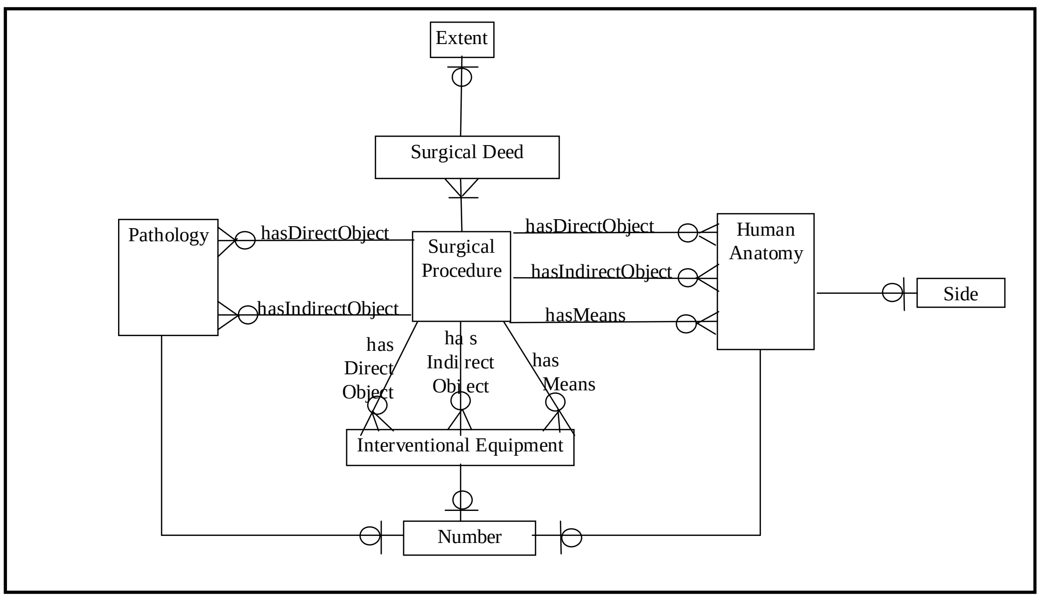


Figure 6b : Basic GALEN scheme for Surgical Procedures

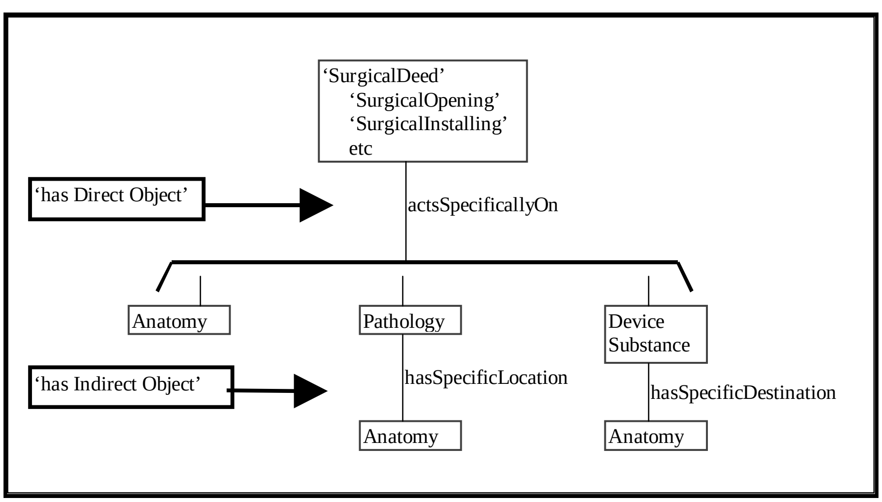
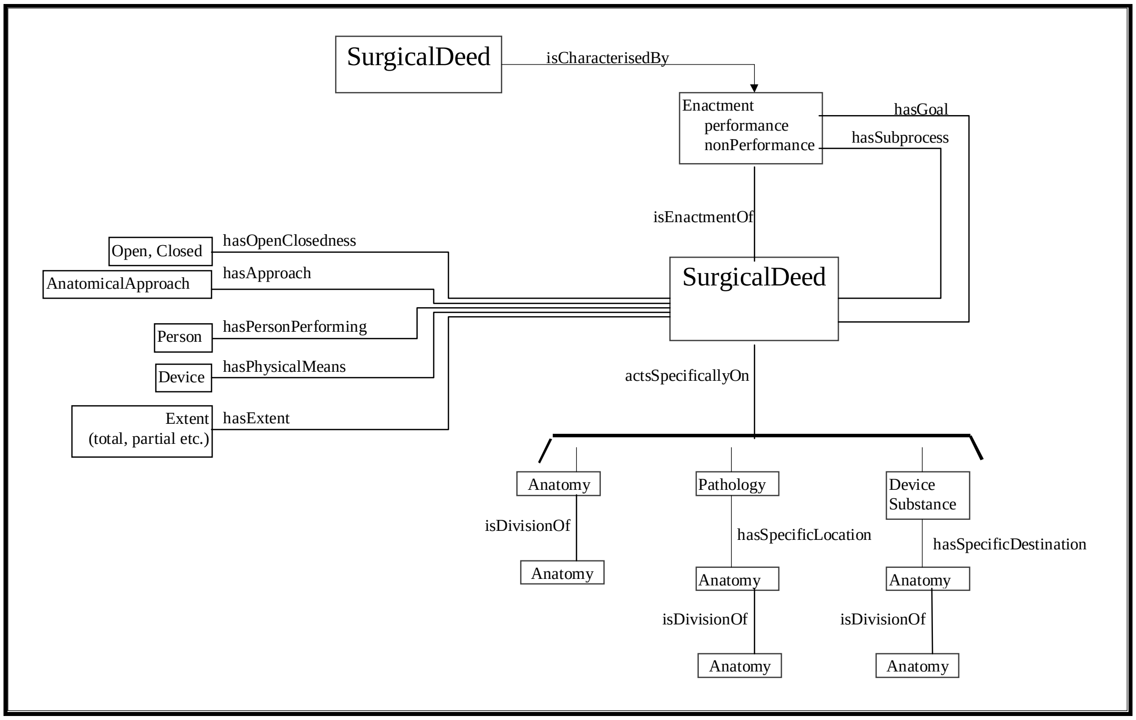


Figure 6c : Extended GALEN scheme for Surgical Procedures



3.5.2 Normalisation and normative statements

The issue of pragmatic normalisation, discussed above, is particularly serious for procedures. Three areas gave particular difficulty:

a) Deeds and intentions: A scheme of levels of intention from simple deeds such as ‘cutting’, to major acts such as ‘removing’ to physiological goals such as ‘revascularisation’ to clinical aims such as ‘curing’ or ‘palliating’ is elaborating by Rossi Mori in [61]. GALEN chose the second level as the primary one for all surgical procedures – *e.g.* “insertion of pins to fixate” rather than “fixation by insertion of pins”.

b) Approach: in descriptions ‘approach’ varies from simple statements such as ‘open/closed’ to complex anatomical routes to entire subprocedure descriptions. The GALEN intermediate representation corresponds to Figure 6c. Internally, approach is normally expanded into a subprocedure.

c) Extent: Extent covers a series of different dimensions – total/partial/…, radical/local/…, These are lumped together in the intermediate representation. Several schemes have been used to separate them internally; none entirely satisfactory. (One of the advantages of the intermediate representation is that such changes can be made easily.)

Whether or not to include ‘normative’ statements in GALEN’s work on procedure classifications has been left entirely to the clinical experts who are asked to produce a ‘paraphrase’ of the procedure rubric which completely represents their understanding which they then formalise in the intermediate representation – see [27].

The key transitive properties of the attributes involved are:

a) *actsOn* and *hasLocation* are transitive

b) *hasSubprocess* is transitive

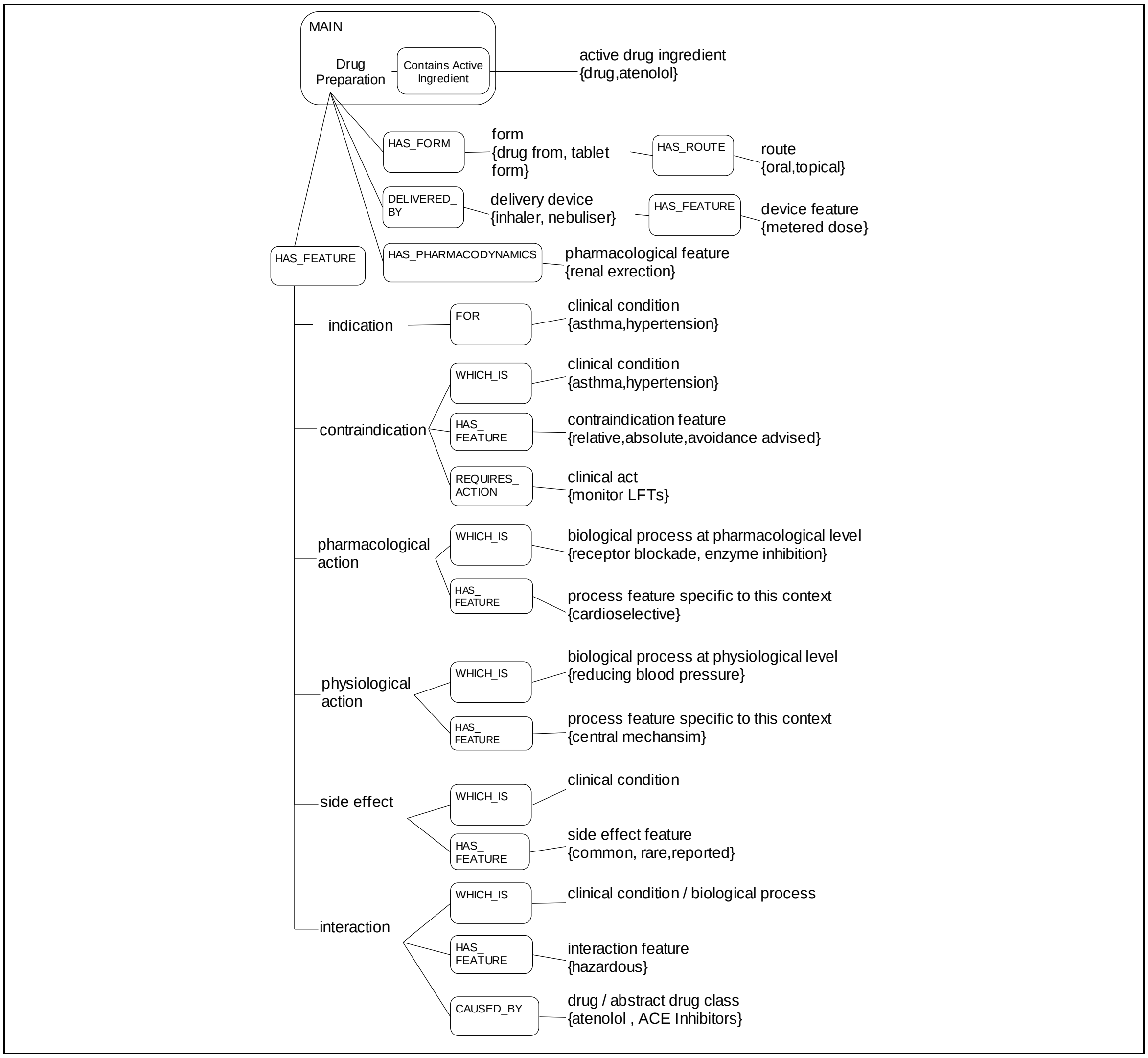
c) anything which *actsOn* a lesion or device located in something *actsOn* that thing.

d) anything which *actsOn* a *division* of something, acts on the whole.

3.6 Drugs

The GALEN drug ontology is now a major focus for development. It is outlined in [36]. The most important fact of the drug ontology is that it is motivated, above all, by linking medical records to decision support and specific difficulties encountered in doing so with traditional classification schemes. The focus of the drug ontology is on the principle of orthogonal taxonomies – representing separately chemical structure, pharmacological (molecular/biochemical) action, physiological action, use, effects, contraindications and side effects. The summary diagram from [36] is presented in Figure 7.

Figure 7: Summary of GALEN Drug Ontology



4. Current Status

The GALEN Common reference model is now in use for development of surgical procedure models by the French, Dutch, and UK governments for development of procedure models and drug ontologies. It is now available on an open source license from www.OpenGALEN.org. The Common Reference Model itself now contains over 8000 concepts, and has been used to 'dissect' and co-ordinate over 15,000 surgical procedure definitions from nine national coding centres. It is the basis of one commercial clinical user interface and of ongoing development within the Synex project http://www.gesi.it/synex/gesi.htm.

Although the ontology is being continuously refined, the main features reported in this paper have been stable for some time. The rate of change stimulated by the emergence of *Open*GALEN remains to be seen.

5. Discussion

The first part of this paper presents a set of major principles for the construction of clinical ontologies; the second illustrates how those principles have been motivated by GALEN's experience and carried through in the development of the GALEN ontology – in some detail with respect to anatomy, and in outline with respect to other areas. The number and complexity of issues illustrates the magnitude of the tasks of developing a re-usable clinial ontology.

In making the decisions illustrated in this paper, GALEN has attempted to adhere to two high level criteria, that choices be motivated by evidence that either:

a) They affected the classification in concepts in ways that were significant to potential applications
b) They were needed so that they could be understood and maintained by the knowledge engineers developing the system.

The limitations of logical representations mean that many technically necessary distinctions technically are not intuitive to users or developers. In particular, the decision as to which information to include in the ontology itself and which information to treat as default or 'extrinsic' information or part of the linguistic and pragmatic layers of the architecture have no obvious parallel in clinicians' intuitions about medicine or about what is 'terminological'. Other constructions such as the use of the feature *Level* to accommodate all possible combinations of absolute magnitude, expectation, change, and gradient are cumbersome in the usual case. Some distinctions are nearly arbitrary, such as whether to treat lesions or the processes which cause those lesions as primary.

The need for distinctions with are necessary for correct performance but which must be made on technical grounds which are counter-intuitive for users and developers alike is a prime motivation for GALEN's use of 'intermediate representations' for knowledge acquisition [23, 28]. It reinforces GALEN's position that, as set out in Table 1, the ontology is only one element in a comprehensive terminology system for clinical information systems.

Informal comparisons with the published material from the Digital anatomist project and its web site (http://www1.biostr.washington.edu/DigitalAnatomist.html) suggest a high degree of concurrence in key areas of the anatomic model. Comparisons with the UK Clinical Terms (Read Codes), version 3 for surgical procedures have been reported elsewhere [57]. The success of the preliminary phases of the development of the GALEN drug ontology as part of the UK Prodigy Project on developing prescribing guidelines for chronic diseases in primary care [36, 62] is providing strong evidence for the efficacy of GALEN's principles, particularly the principle of orthogonal taxonomies, in allowing re-use in a key clinical application where traditional approaches had encountered serious difficulties. Comparison with the developments in SNOMED-RT [2, 3] should soon be possible.

At least seven major issues remain outstanding:

a) Improved pragmatic normalisation where logical criteria do not suffice .
b) Improved handling of congenital abnormalities
c) Improved handling of 'normative' statements
d) Improved handling of other patterns for the interaction of transitive semantic links
e) Improved handling of adjacency and other spatial and temporal notions
f) How to best take advantage of the improvements in description logic technology now becoming available [58, 63].

Issue a) is primarily an issue of tools and environments; issues b) and c) concern primarily the ontology itself; while issues d) and e)concern the interaction of the ontology with potential extensions to the formalism in which it is represented.

The material in the appendices of this paper provide a starting point for examining GALEN's ontology. The complete ontology is available through the *Open*GALEN Foundation at www.opengalen.org. Ultimately, evaluation of clinical ontologies remains an empirical task based on experimentation with different applications – whether the ontologies, along with the other elements of a terminology system listed in Table 1, can be used effectively to support clinical information systems.

Acknowledgements

The material in this paper is the distillation of many hours of discussion and many varied contributions from members of the GALEN and GALEN-IN-USE consortia, the PEN&PAD team, and their collaborators. Particular thanks to Anthony Nowlan from whom many of the ideas originated and to Angelo Rossi Mori who provided many of the critiques which forced consideration of key issues – whether successfully or not he and others will have to judge. This work was supported in part by grants from the European Commission, the UK Department of Health, and the UK Medical Research Council.

References

1. Rector, A, Bechhofer, S, Goble, C, Horrocks, I, Nowlan, W, and Solomon, W, *The GRAIL concept modelling language for medical terminology.* Artificial Intelligence in Medicine, 1997. **9**: pp. 139-171.
2. 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.
3. Spackman, KA and Campbell, KE, *Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies.* Jounral of the American Medical Informatics Association, 1998(1998 Fall Syposium special issue): pp. 740-744.

4. Campbell, KE, Tuttle, MS, and Spackman, KA, *A "lexically-suggested logical closure" metric for medical terminology maturity.* Journal of the American Medical Informatics Association, 1998(Fall Symposium Special Issue): pp. 785-789.

5. Evans, DA, Cimino, J, Hersh, WR, Huff, SM, Bell, DS, and The Canon Group, *Position statement: Towards a medical concept representation language.* Journal of the American Medical Informatics Association, 1994. **1**(3): pp. 207-217.

6. Cimino, J, *Controlled Medical Vocabulary Construction: Methods from the Canon Group.* Journal of the American Medical Informatics Association, 1994. **1**(3): pp. 296-197.

7. Tuttle, MS, *The position of the canon group: A reality check.* Journal of the American Medical Informatics Association, 1994. **1**(3): pp. 298-299.

8. 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.

9. 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.

10. 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.

11. Musen, M, *Modern architectures for intelligent systems: reusable ontologies and problem-solving methods.* Journal of the American Medical Informatics Association, 1998(Symposium supplement): pp. 46-54.

12. Erikson, H, Shahar, Y, Tu, S, and Puerta, A, *Task modelling with reusable problem-solving methods.* Artificial Intelligence, 1995. **79**: pp. 293-326.

13. Tu, S, Eriksson, H, Gennari, J, Shahar, Y, and Musen, M, *Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools - application of PROTEGE-II to protocol-based decision-support.* Artificial Intelligence in Medicine, 1995. **7**: pp. 257-289.

14. Goble, C, Haw, C, and Bechhofer, S. *Describing and classifying multimedia using the description logic GRAIL.* in *Proceedings of the SPIE Conference on Storage and Retrieval of Still Image and Vidio IV, SPIE vol 2670*. 1995. San Jose: pp. 132-143.

15. Guarino, N, *Understanding, building and using ontologies.* International Journal of Human Computer Studies, 1997. **46**: pp. 293-310.

16. Borgida, A, *Description logics in data management.* IEEE Transactions on Knowledge and Data Engineering, 1995. **7**(5): pp. 671-682.

17. Lenat, DB and Guha, RV, *Building Large Knowledge-Based Systems: Representation and inference in the Cyc Project*. 1989, Reading, MA: Addison-Wesley. 372.

18. Fikes, R, Farquhar, A, and Pratt, W. *Information brokers: Gathering information from heterogeneous information sources.* in *Ninth Florida Artificial Intelligence Research Symposium (FLAIRS-96)*. 1996. Key West, Florida: pp. 192-197.

19. Motta, E, Buckingham-Shum, S, and Domingue, J. *Ontology-Driven Document Enrichment: Principles and Case Studies*. in *Knowledge Acquisition Workshop*. 1999. Banf, Canada: pp. http://sern.ucalgary.ca/ksi/kaw/KAW99/papers.html.

20. Fensel, D, Decker, S, Erdmann, M, and Studer, R. *Ontobroker: How to enable intelligent Access to the WWW.* 1999: pp. www.aifb.uni-karlsruhe.de/WBS/dfe/iii99.html.

21. Hardiker, NR and Rector, AL, *Modeling nursing terminology using the GRAIL representation language.* Journal of the American Medical Informatics Association, 1998. **5**: pp. 120-128.

22. Rogers, J and Rector, A. *The GALEN ontology.* in *Medical Informatics Europe (MIE 96)*. 1996. Copenhagen: IOS Press: pp. 174-178.

23. Rector, AL, *Clinical Terminology: Why is it so hard?* Methods of Information in Medicine, 1999. **38**: pp. (in press).

24. Rector, A, *Thesauri and formal classifications: Terminologies for people and machines.* Methods of Information in Medicine, 1998. **37**(4-5): pp. 501-509.

25. 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.

26. Cimino, J, *Desiderata for controlled medical vocabularies in the twenty-first century.* Methods of Information in Medicine, 1998. **37**(4-5): pp. 394-403.

27. 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.

28. 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.

29. Ceusters, W, *The distinction between linguistic and conceptual semantics in medical terminology and its implications for NLP-based knowledge acquisition.* Methods of Information in Medicine, 1998. **37**(327-333).

30. Rassinoux, A-M, *Modeling just the important and relevant concepts in medical language understanding.* Methods of Information in Medicine, 1998(37): pp. 361-372.

31. Zanstra, P, Claassen, W, and Haring, Evd, *A natural language generator in Rois.* LINKS, 1997(2): pp. http://ehm.kun.nl/mi/project/rois/newsltr/Links2/Paper4/paper4.htm.

32. 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.

33. Ceusters, W, Rogers, J, Consorti, F, and Rossi-Mori, A, *Syntactic-semantic tagging as a mediator between linguistic representations and formal modals: an exercise in linking SNOMED to GALEN.* Artificial Intelligence in Medicine, 1999. **15**(1): pp. 5--24.

34. Nowlan, W, Rector, A, Rush, T, and Solomon, W. *From Terminology to Terminology Services.* in *18th Annual Symposium on Computer Applications in Medical Care (SCAMC-94)*. 1994. Washington, DC: pp. 150-154.

35. Rector, A, Zanstra, P, Solomon, W, and The GALEN Consortium, *GALEN: Terminology Services for Clinical Information Systems*, in *Health in the New Communications Age*, M. Laires, M. Ladeira, and J. Christensen, Editors. 1995, IOS Press: Amsterdam. p. 90-100.

36. Solomon, DS, Wroe, C, Rogers, JE, and Rector, A, *A reference terminology for drugs.* Journal of the American Medical Informatics Assciation, 1999(1999 Special Conference Issue).

37. Brachman, R and Levesque, H. *The tractability of subsumption in frame-based description languages.* in *AAAI-84*. 1984: Morgan Kaufman: pp. 34-37.

38. 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.

39. Rossi Mori, A and Consorti, F, *Structures of Clinical Information in Patient Records.* Journal of the American Medical Informatics Association, 1999(Fall Symposium Special Special Issue): pp. in press.

40. Rossi Mori, A and Consorti, F. *Structuring Clinical Information in Healthcare Records.* in *MIE '99*. 1999. Lubljana: pp. (in press).

41. Brachman, RJ, McGuinness, DL, Patel-Schneider, PF, Resnick, LA, and Borgida, A, *Living with Classic: When and how to use a KL-ONE-like language*, in *Principles of Semantic Networks: Explorations in the representation of knowledge*, J. Sowa, Editor. 1991, Morgan Kaufmann: San Mateo, CA. p. 401-456.

42. 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.

43. Grosso, WE, Eriksson, H, Fergerson, RW, Gennari, JH, Tu, SW, and Musen, MA. *Knowledge modelling at the millenium (The design and evolution of Protege-2000).* in *Knowledge acqusition workshop (KAW-99)*. 1999. Banf, Canada.

44. O'Neil, M, Payne, C, and Read, J, *Read Codes Version 3: A user led terminology.* Methods of Information in Medicine, 1995. **34**: pp. 187-192.

45. Sowa, J, *Knowledge Representation*. 1999: Morgan Kaufmann.

46. Bateman, JA, *Upper modelling: a general organization of knowledge for natural language processing.*, . 1989, USC/Information Sciences Institute:

47. Guha, R and Lenat, D, *Enabling agents to work together.* Communications of the ACM, 1994. **37**: pp. 127-142.

48. Schank, RC and Abelson, R, P, *Scripts, Plans, Goals, and Understanding*. 1977, Hilsdale, NJ: Lawrence Erlbaum Associates.

49. Etherington, D, *Formalising nonmonotonic reasoning systems.* Artificial Intelligence, 1987. **31**: pp. 41-85.

50. Touretzky, D, *The Mathematics of Inheritance Systems.* 1986, Los Altos, CA: Morgan Kaufmann.

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. Padgham, L and Lambrix, P. *A Framework for Part-of Hierarchies in Terminological Logics.* in *KR-94*. 1994: pp. 485-96.

54. Artale, A, Franconi, E, and Guarino, N. *Open problems for part-whole relations.* in *International Workshop on Descripition Logics*. 1996. Boston, MA: pp. http://www.cs.man.ac.uk/franconi-bin/retrieve/links/franconi-papers/part-papers/part-dl96.dvi.gz.

55. Artale, A, Franconi, E, and Pazzi, L, *Part-whole relations in object-centered systems: An overview.* Data and Knowledge Engineering, 1996. **20**: pp. 347-383.

56. 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).

57. 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.

58. 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).

59. Baader, F and Hollunder, B. *A terminological knowledge representation system with complete inference algorithm.* in *Proceedings of the Workshop on Processing Declarative Knowledge, PDK0-91 (Lecture Notes in Artificial Intelligence #567)*. 1991: Springer-Verlag: pp. 67-86.

60. Shahar, Y, Tu, S, Das, A, and MA, M. *A problem-solving architecture for managing temporal data and their abstractions*. in *Workshop on Implementing Temporal Reasoning, AAAI-92*. 1992. San Jose, CA: pp. 141-52.

61. Rossi Mori, A, Gangemi, A, Steve, G, Consorti, F, and Galeazzi, E. *An ontological analysis of surgical deeds*. in *Artificial Intelligence in Medicine Europe (AIME-97)*. 1997. Grenoble, France.

62. Purves, I, Sowerby, M, Beaumont, R, and Sugden, R. *An aristocrat, a precocious child and a sacred cow. Developing general practice clinical computer systems: a collaboration between general practitioners, system suppliers and the NHS Executive - lessons from the Prodigy methodology*. in *Annual Conference of the Primary Health Care Subgroup of the British Computer Society*. 1996. Cambridge, England: pp. 51-73.

63. Horrocks, I. *The CAMELOT Web Page*. 1998: http://www.cs.man.ac.uk/~horrocks/Camelot/

Appendix A: Further examples of the GALEN ontology

Figure A1: Taxonomy of major elementary Categories
(levels of indenting denote the subsumption hierarchy)

Entity	**Example**
GeneralisedProcess	
SpecificProcess	
BiologicalProcess	
BodyProcess	Peristalsis, Breathing, Clotting
Behaviour	VolitionalAct, ClinicalAct
NonBiologicalProcess	
PhysicalProcess	Irradiation
ChemicalProcess	Histological Staining
GenericProcess	Transport, Opening, Closing
GeneralisedStructure	
AbstractStructure	
PsychosocialConstruct	Clinic, Hospital
LogicalStructure	Protocol, Plan
PhysicalStructure	
LinearStructure	Route
PlanarStructure	Triangle, Square
SolidStructure	
MicroscopicStructure	Cell, Microorganism
InertSolidStructure	Building, Device
OrganicSolidStructure	
BodyStructure	
BodyPart	Heart,Leg, Head, Femur, AdrenalGland, Sacrum, PisiformBone,
GenericBodyStructure	Cusp, Horn, Promontory, Artery, Lump, Bursa, Orifice, Rim, Ridge
Organism	Bacterium. Protozoan, Virus, Fungus, Dog, Bird
SolidRegion	PieceOfLiver.
Space	Cavity, Potential Space, PathologicalCavity
GeneralisedSubstance	
Energy	Radiation, SoundEnergy
Substance	
BodySubstance	
Tissue	MuscleTissue, BoneTissue, Mucosa, Endothelium
NAMEDBodySubstance	Urine, Bile , Sputum, Vomit, Sputum .
ChemicalSubstance	Drugs, Sodium .
NAMEDSubstance	Air, Wood
ModifierConcept	
Aspect	
Feature	Sex, Chronicity, Shape, Malignancy, Topology, Colour, Permeability .
State	male/female, acute/chronic, round/square, permeable/impermeable .
Selector	leftSelection/rightSelection, medialSelection/lateralSelection
Status	pathological/pysiological normal/nonNormal
Collection	Polyps (as opposed to polyp)
Modality	FamilyHistory, PreviousHistory, presence/absence
Role	DrugRole, HormoneRole, PatientRole
Unit	second, metre, kilogram
LevelOfSpecification	uniquelySpecified

Figure A2 : Taxonomy of major elementary categories for anatomical/morphology model

GRAIL Category	Examples of categories below this point
BodyStructure	
BodyRegion	segment of intestine, body layer, perianal region
BodyConnection	
Fistula	arterio-venous fistula, cutaneous fistula, stoma
AnastomosisStructure	
ShuntStructure	arterio-venous shunt structure
BypassStructure	coronary artery bypass structure
ArbitraryBodyConstruct	
GenericBodyStructure	fissure, edge, flap, cord, pouch, hair
GenericBodySurfaceStructure	
GenericInternalStructure	valve, duct, artery, vein, bursa, apex, lobe, gland
BodyPart	
NAMEDSurfaceBodyPart	
ExternalOrgan	skin as organ
MajorBodyDivision	head, neck, trunk, upper extremity
SurfaceBodyLandmark	waist, bregma, linea alba
SurfaceOpening	anus, mouth, introitus
BodyJunctionalBodyPart	axilla, groin, shoulder, hip
NAMEDHeadSurfaceBodyPart	cheek, chin, ear, face, forehead, lip
…	
NAMEDSurfaceSubpart	heel, shin, calf, palm, occiput
NAMEDEmbryonicBodyPart	paramesonephricus, neural tube
NAMEDInternalBodyPart	
InternalOrgan	gall bladder, heart, kidney, larynx
NAMEDGITractBodyPart	esophagus, anal canal, colon, gastric fundus
…	
NAMEDInternalBodySubPart	
NAMEDBoneDivisions	olecranon, styloid process, condyle, shaft, ramus

Figure A3: Summary of major partitive attributes

HasDivision	Does it divide into similar pieces ?	
hasSurfaceDivision	- in two dimensions ?	Palm hasSurfaceDivision ThenarEminence Hand hasSurfaceDivision Palm
hasSolidRegion	- in three dimensions ?	
hasLayer	When the structure is divided, is there still a layer in each division?	Stomach hasLayer Mucosa Bone hasLayer Cortex
hasBlindPouchDivision	Specific to the appendix vermiformix etc.	Appendix isBlindPouchDivisionOf Caecum Alveolus isBlindPouchDivisionOf Bronchus
hasSolidDivision		Stomach hasSolidDivision GastricFundus Heart hasSolidDivision CardiacSeptum
hasLinearDivision	Can it be divided into segments? Does obstruction of a division obstruct the whole?	Intestine hasLinearDivision Jejunum LongBone hasLinearDivision Epiphysis
hasStructuralComponent	When a structure is divided, does the component (usually) reside in one division?	KneeJoint hasStructuralComponent Meniscus Neck hasStructuralComponent ThyroidGland
isMadeOf	For mass items — liquids, tissues, etc.	Meniscus isMadeOf Fibrocartilage Thrombus isMadeOf ClottedBlood

Figure A4: Summary of major structural attributes

DelimitingAttribute		
ContainmentAttribute	Does one object contains another ?	
hasMixedThroughout	is the container solid or a liquid	Liver hasMixedThroughout Metastases
boundSpace	Is the contained structure a space ?	JointCapsule boundsSpace Joint
contains	Is the container hollow ?	AbdominalCavity contains Liver Pericardium contains Heart
LinearContainmentAttribute	Is one of the structures a linear structure ?	
PassesThrough	does the line pass through a structure ?	TransdermalRoute passesThrough Skin
hasBranch	is one structure a branch of the other ?	Trachea hasBranch MainBronchus
hasColinearityWith	Is one linear structure colinear with the other ?	TransvaginalRoute hasColinearityWith Vagina
hasParallelToIt	Is one linear structure parallel to the other ?	ParamedianIncision hasParallelToIt MidlineBodyLine
serves	Is one structure a blood vessel or nerve ?	PhrenicNerve serves Diaphragm
connects	Does one structure connect one or more other structures ?	MitralValve connects LeftHeartVentricle
hasProximity	Is one structure near another ?	PerianalTissue hasProximity Anus

Figure A5 : Causal Attributes

- isCausallyLinkedTo
  - isAnswerOf
    - isSpecificAnswerOf
  - isAssociatedWith
    - isSpecificallyAssociatedWith
  - isConsequenceOf
    - isComplicationOf
      - isSpecificComplicationOf
        - isSpecificImmediateComplicationOf
        - isSpecificLateComplicationOf
    - isImmediateConsequenceOf
      - isSpecificImmediateConsequenceOf
        - hasUniqueAssociatedProcess
        - isSpecificallyTriggeredBy
      - isUniqueImmediateConsequenceOf
    - isLateConsequenceOf
    - isSpecificConsequenceOf

Figure A6: Summary of major selective and descriptive attributes.

(The A≥B means that A subsumes B; some value sets shown here are arranged hierarchically in the model.)

Selector Attributes	**Permitted range of values**
hasLeftRightSelector	[leftSelection rightSelection]
hasVisceralParietalSelector	[visceralSelection parietalSelection]
hasMedialLateralSelector	[medialSelection lateralSelection]
hasAnteriorPosteriorSelector	[anteriorSelection posteriorSelection]
hasSuperiorInferiorSelector	[superiorSelection inferiorSelection]
hasProximalDistalSelector	[proximalSelection centralSelection distalSelection]
hasUpperLowerSelector	[lowerSelection middleSelection upperSelection]
hasInnerOuterSelector	[innerSelection outerSelection]
hasDorsalVentralSelector	[dorsalSelection ventralSelection]
hasGreaterLesserSelector	[greaterSelection lesserSelection]
hasBilateralUnilateralSelector	[bilateralSelection unilateralSelection ≥ [ipsilateralSelection contralateralSelection]]
Descriptive Attributes	**Permitted range of values**
hasTopology	[solid, hollow ³ [tubular, closedHollow, actually hollow, bilayered]]
hasSurfaceVisibility	[surfaceVisible, internal]
hasSeverity	[mild, moderate, severe]
hasCountability	[mass, discrete]
hasShape	[laminar, linear, elongatedSolid, toroidal, spherical …]
isPairedOrUnpaired	[unpaired, atLeastPaired ³ [moreThanPaired exactlyPaired ³[leftRightPaired ³ [mirrorImaged]]]]

Figure A7 RoleDesignatingAttributes and Roles

- RoleDesignatingAttribute
  - hasHCMRole
  - playsClinicalRole
  - playsConsultationRole
  - playsNumericalRole
  - playsPathologicalRole
  - playsPhysiologicalRole
  - playsRelativeImportanceRole
  - playsSocialRole

- Role
  - ClinicalRole
    - ConsultationRole
      - DiagnosisRole
      - SignRole
      - SymptomRole
    - SurgicalRole
      - MajorSurgicalRole
      - MinorSurgicalRole
  - NumericalRole
    - MaximumRole
    - MinimumRole
  - PathologicalRole
    - AddictiveRole
    - AllergicRole
    - InfectiveRole
    - IntoxicationRole
  - PhysiologicalRole
    - AntigenRole
    - BiochemicalRole
      - AntigenicRole
      - DrugAdditiveRole
      - DrugRol e
      - EnzymeRole
      - HormoneRole
        - SexHormoneRole
    - FoodRole
      - NonNutritiveFoodRole
      - NutritiveFoodRole
    - FunctionalRole
      - EndArteryRole
      - ExerciseRole
      - GreatVesselRole
      - …
    - SampleRole
      - BiopsyRole
      - SpecimenRole
    - StructuralRole
      - PhysicalBarrierRole
      - PhysicalProtectionRole
      - PhysicalSupportRole
      - PresentingPartRole
  - SocialRole
    - HobbyRole
      - …
    - MemberRole
    - OccupationRole
      - AdministratorRole
      - BakerRole
      - …
      - DieticianRole
      - DoctorRole
        - SurgeonRole
      - FarmerRole
      - t…
      - NurseRole
        - TheatreNurseRole
      - …
    - PatientRole
  - SubsidiaryRole

Figure A7: Major Pathological Processes

- NAMEDPathologicalProcess
  - VomitingProcess
  - HaemorrhagingProcess
  - …
  - AcquiredPathologicalProcess
    - TraumaticProcess
      - ForeignBodyAcquisitionProcess
      - …
    - AbnormalGrowthProcess
      - NeoplasticProcess
      - AbnormalCellMassGrowthProcess
      - …
    - IatrogenicLesionProcess
    - IdiopathicLesionProcess
    - IschaemicProcess
- InflammatoryProcess
    - UlceratingOrErosionProcess
    - CellulitisProcess
    - …
- InfectionProcess
  - HypersensitivityProcess
  - PerforatingProcess
    - HerniationProcess
    - RupturingProcess
    - DiverticulumProcess
  - StrictionProcess
  - DilatatingProcess
  - DegenerativeProcess
    - DegenerativeAtrophicProcess
    - …
  - IntussusceptionProcess
- CongenitalPathologicalProcess

Figure A8: Summary of Transitive Attributes in GALEN

Note: transitivity is NOT inherited. *A transitive* and *A* subsumes *B* does not imply *B transitive*.

- Attribute / InverseAttribute
  - HasDivision / IsDivisionOf
    - hasSurfaceDivision / isSurfaceDivisionOf
    - hasSolidDivision / isSolidDivisionOf
    - hasLayer / isLayerOf
    - hasStructuralComponent / isStructuralComponentOf
    - hasFunctionalComponent / isFunctionalComponentOf
  - isMadeOf / makesUp
  - hasSubprocess / isSubprocessOf
  - ContainmentAttribute / InverseContainmentAttribute
  - LocativeAttribute / InverseLocativeAttribute
  - hasParallelToIt / isParellelTo
  - isPositionedSuperiorTo / isPositionedInferiorTo
  - actsOn / isActedOnBy
  - isConsequenceOf / hasConsequence
    - sComplicationOf / hasComplication
  - isCharacterisedBy / characterises
  - occursDuring / hasEventOccuringDuring
  - isFollowedBy / follows

Figure A9: Summary of major specialisedBy axioms in GALEN

Definitions:

If A_1 *specialisedBy* A_2 then

X which A_1 *Z* subsumes *X which* A_1 *(Y which* A_2 *Z)*

and

Z which inv A_1 *X* subsumes *Z which inv* A_2 *(Y which inv* A_1 *X)*

or in terms of composition operators more familiar in description logic notations

If A_1 *specialisedBy* A_2 then

A_1 subsumes $A_1 \circ A_2$

and hence

inv A_1 subsumes *inv* $A_2 \circ$ *inv* A_1

Properties:

A_2 subsumes B_2 and A_1 specialisedBy A_2, then A_1 specialisedBy B_2.

A_1 specialisedBy A_2 does NOT imply A_2 specialisedBy A_2, but does not forbid it.

Attribute	specialisedBy
LocativeAttribute (involves)	isDivisionOf, isMadeOf, isMultipleOf, isPhysicalMeansOf, isSpacedDefinedBy
isStructuralComponentOf	isSurfaceDivisionOf, isSolidDivisionOf, isSpaceDefinedBy
serves	isDivisionOf
isFunctionOf	isFunctionalComponentOf
hasFunction	isSubProcessOf
actsOn	isFunctionOf, hasLocation, isDivisionOf, isMadeOf
isActedOnBy	isImmediateConsequenceOf, isSubprocessOf
osAssociatedWith	isImmediateConsequenceOf
isCharacterisedBy	isExistenceOf

Appendix B: Links to related papers on line

Full bibliography at http://www.cs.man.ac.uk/mig/giu/biblio.html

Relevant references on line:

Rector, A.L., Rogers J.E, Pole P.M. The GALEN High Level Ontology (postscript) (RTF). presented at MIE 96, Copenhagen http://www.cs.man.ac.uk/mig/ftp/pub/papers/MIE96.ps

Elena Galeazzi, Angelo Rossi Mori, Fabrizio Consorti, Anna Errera A cooperative methodology to build conceptual models in medicine In Medical Informatics Europe '97 pp280-284 (Thessaloniki, Greece) IOS Press Vol 43 ISSN:0926-9630 http://www.cs.man.ac.uk/mig/giu/papers/mie97/mie97cnr.rtf

Rogers J.E, WD Solomon, Rector, A.L., Pole P.M., P Zanstra, E van der Haring, (1997) Rubrics to Dissections to GRAIL to Classifications (RTF) In Medical Informatics Europe '97 pp241-245 (Thessaloniki, Greece) IOS Press Vol 43 ISSN:0926-9630 http://www.cs.man.ac.uk/mig/giu/papers/mie97/mie97vum.rtf

J-M Rodrigues, B Trombert-Paviot, R Baud, J Wagner er al Galen-In-Use : An E.U Project applied to the development of a new national coding system for surgical procedures : NCAM In Medical Informatics Europe '97 pp897-901 (Thessaloniki, Greece) IOS Press Vol 43 ISSN:0926-9630 http://www.cs.man.ac.uk/mig/giu/papers/mie97/mie97use.rtf