Ontological and Practical Issues in Using a Description Logic to Represent Medical Concept Systems: Experience from GALEN

Alan Rector and Jeremy Rogers

School of Computer Science, University of Manchester, Manchester M13 9PL, UK
rector@cs.man.ac.uk, rogers@cs.man.ac.uk

Abstract. GALEN seeks to provide re-usable terminology resources for clinical systems. The heart of GALEN is the Common Reference Model (CRM) formulated in a specialised description logic. The CRM is based on a set of principles that have evolved over the period of the project and illustrate key issues to be addressed by any large medical ontology. The principles on which the CRM is based are discussed followed by a more detailed look at the actual mechanisms employed. Finally the structure is compared with other biomedical ontologies in use or proposed.

1 Introduction

1.1 Background

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 (CRM), formulated in a specialised description Logic, GRAIL [46] . Since GALEN's inception there have been several other major efforts at medical "ontologies", the most important being SNOMED-CT[1] which has been made widely available in the United States via licensing by the National Library of Medicine and in the UK via licensing to the National Health Service [75].

Likewise, since GALEN's inception, "ontologies" have come to be much more widely studied in relation both to information systems theoretically (*e.g.* [20] [69]), practically (*e.g.* [5, 18, 81]) , in biomedical applications generally (*e.g.* [23, 70]) and in specific areas such as anatomy [32, 63]. Indeed, a track on "ontologies" is a feature of many conferences on the Semantic Web and database design in biohealth informatics. GALEN itself drew heavily on the pioneering work of the CANON group [11, 16, 79] and on ideas from early phases of the Cyc project [31].

GALEN has been used, amongst other activities, for the development of the French national classification of surgical procedures CCAM [57], as part of the procedure for revising the Dutch classification of procedures, in the development of a drug ontology in the UK [72, 87] and in associated work "untangling" forms and routes of drug administration as part of a collaboration with HL7 [86]. Two independent studies have examined the issues in reconciling GALEN's modelling of anatomy with that of

[1] http:/www.snomed.org

P. Barahona et al. (Eds.): Reasoning Web 2006, LNCS 4126, pp. 197 – 231, 2006.

the Digital Anatomist Foundational Model of Anatomy [35, 36, 88-90]. GALEN has also given rise to a methodology for normalising ontologies to promote modularisation [44].

This paper presents a unified approach to the principles and details of the GALEN Common Reference Model (CRM), previously partly described in [26, 58, 59]. GALEN's CRM is one of four models at the core of an overall architecture for use, and re-use, of clinical terminology [49, 51, 56]. A discussion of broader issues and the relation to Cimino's desiderata for clinical terminologies [12] can be found in [52]. A discussion of the use of the ontology in representing pharmaceutical information can be found in [72, 87]. The discussion section of this paper reassesses some of the decisions in the GALEN CRM in terms of developments since its inception in the early 1990s and includes a brief comparison with Welty and Guarino's Ontoclean/Dolce [18, 19, 83] and Smith's Basic Formal Ontology (BFO) [69, 70].

1.2 GALEN's Aims and Criteria for Success

The overall aim of the GALEN terminology resources is to support clinical information systems. For individual patients, it aims to allow clinical information to be recorded faithfully in their electronic record, and then abstracted from it. Such abstraction supports re-organisation or filtering to provide a clearer view of the patient, and linkage to knowledge resources such as decision support, bibliographic, and general web-based information systems. For populations of such patients it supports aggregation for secondary re-use in management, research, and administrative contexts. Abstraction, re-organisation, and re-use are fundamentally dependent on classification, and therefore the primary technical criterion for the GALEN ontology is: *correct and complete classification of its definitions and descriptions.*

More generally, we can describe any ontology in terms of:

1. *Expressiveness* – the ability to represent formally the notions required by its users; for medical ontologies this means all relevant symptoms, diseases, procedures, etc.
2. *Classification* – the ability to infer the correct classification (indexing) of the expressions represented, a) soundly, and b) completely, where by "soundness" we mean that all inferences made are correct, and by "completeness" that all possible sound inferences are made.
3. *Parsimony* – GALEN was specifically designed for use as a "post-coordinated system", in which the classification of new expressions is inferred and dynamically maintained *post hoc*. This avoids the combinatorial explosion inevitable with pre-coordinated systems, in which all legitimate expressions must be pre-enumerated and classified *pre hoc*. An explicit goal of post-coordinated systems is to obtain maximum expressiveness from a finite and limited range of basic notions.

Achieving these goals, however, still requires greater complexity than clinical authors can be expected to cope with. The GALEN ontology is, therefore, designed as an internal 'assembly language', rarely to be seen directly by users or even by most software developers. Intuitive, user-oriented presentation is handled separately through 'intermediate representations' described elsewhere [56, 60].

2 Rationale for the GALEN Common Reference Model

2.1 Basic Principles

2.1.1 'Logical Approximations'

Any logical model for knowledge representation is at best an approximation of the relevant concepts as used in human language and thought. A "logical approximation" may seem an oxymoron, but logical models of any kind behave very differently from language or our internal conceptualisation. Thought and language are typically dependent on context in a fluid manner that eludes the rigidity of logical representation for at least three reasons:

1. Logic, at least standard first order logic and description logics, are "two-valued" – they deal only in truth and falsehood. 'Shades of grey', or probabilities, are not supported.
2. There are well known trade-offs between expressiveness and computational tractability in computational logical systems [6, 14].
3. Reality is fractal – no matter how much detail a model represents, it is always possible to represent more. Hence every formal representation must make choices of what to represent.

2.1.2 'Linguistic Approximations'

Since any ontology is an approximation, the labels attached to representations internally in the ontology are necessarily also at best approximations. Arguments such as "Is the hand still a division of the upper extremity if it has been amputated?" or "Is there a difference between an 'act' and a 'deed'?" rarely affect the utility of the ontology for the intended applications. When arguments over the labelling of representations occur, the GALEN team asks two questions:

1. Does the representation represent *some* entity that most users or authors agree to be useful and clearly defined, even if they cannot agree on what it should be called?
2. Is the label seriously misleading? Ambiguous? Does it mean different things to different groups?

With respect to 1), GALEN has usually found agreement on substance to be easier than agreement on the words to describe that substance. Once the two issues are separated, agreement is possible. For example, whether "neoplasm" should mean any new growth or only any specifically malignant new growth was a matter for great debate. There was no debate, however, regarding whether or not separate representations could and should exist for each of "new growth, whether benign or malignant" and "malignant new growth", merely about how they should be named.

With respect to 2), GALEN has found non-understanding to be better than misunderstanding. Internal labels are often deliberately awkward, *e.g.* "*PathologicalPhenomenon*" rather than "disease" or "disorder".

2.1.3 Canonical Forms and "canonization"

Most notions can be represented in more than one logically and/or semantically equivalent form. Although humans recognise such equivalences easily, one such form must be selected as ‘canonical’ [16] if logical computational systems are to be able to manipulate representations and data consistently. GALEN recognises two distinct levels of transformation (“canonization”) between equivalent forms to be dealt with:

1. Logical – *e.g.* to transform “fracture of a long bone located in the femur” to “Fracture of femur”. This is a purely logical operation dependent on the representations of “Fracture”, “Long bone” and “Femur”, where “Femur” is a more specific subclass of ‘Long bone”.
2. Ontological – *e.g.* to transform variants such as “Fixation of femur by means of insertion of pins” and “Insertion of pins to fixate femur” to their preferred form [39]. Such variant forms are not logically equivalent – “Fixations” are not kinds of “Insertions” nor *vice versa* [56]). Such alternatives can be resolved only by metamodel conventions embodying ontological commitments. (See 0.)

2.2 Ontological Issues

2.2.1 Categories, Instances and Natural Kinds

The GALEN Common Reference Model (CRM) contains only “categories” (“classes”)[2] and not “instances”.

Categories can be abstract, such as “phenomenon” or “disease”, general such as “blood dyscrasia” or very specific such as “sugar-free syrup” or “foot”. In principle, however, all categories can be specialised to define new categories which can in turn be further specialised, indefinitely – *e.g.* “sugar-free syrup” to “flavoured sugar-free aspirin syrup”; “foot” to “left foot”, “deformed foot”, “deformed left foot”, etc.

Statements in real world medical records represent statements about “instances”[3] of these categories and, by contrast to categories, can not have kinds or subclasses (can not be “specialised”). It makes no sense to say “a sugar-free kind of *this* tablet of Aspirin” or “a kind of *Alan’s* left foot.”.

Some authors on ontologies identify instances as being entities specialised to the level of detail required for a particular application, e.g Brachman *et al.*’s “Living with Classic” [7]. This approach is fatal to re-use, since, as Brachman *et al.* so elegantly demonstrate, the appropriate level of detail for different applications will almost certainly be different. It is for this reason that the GALEN Common Reference Model contains only categories and no instances.

However, even though it deals only with categories, GALEN must still decide which categories should be “elementary” (“primitive”)[4] and which “composite” (“defined”)[5], *i.e.* defined by expressions made up of other categories. GALEN considers two issues in deciding whether to represent a given entity as elementary or composite:

---

[2] GALEN categories are known variously in other systems as “types”, “classes”, or in Welty & Guarino’s writing “predicates”. The status of what many call “concept” is controversial; we use the word “entity” throughout this paper as a neutral term for either an instance or a category although - since GALEN does not represent instances - “entity” and “category” are for most purposes synonymous.

[3] In some other systems known variously as “individuals”

[4] Also known as “primitive”

[5] Also known as “defined”

1. Whether it is possible to define the category. A definition must give the complete set of all necessary and sufficient criteria for recognising that category. Many important categories defy complete definition by sufficient criteria. Such categories are related to concepts that are often termed "natural kinds" and include most simple notions such as "leg", "tree", "process", "flow", etc. Natural kinds can also occur at a more abstract level. For example, one might be tempted to define "Heart valve" as equivalent to "valve in the heart", and "valve" as a "structure that controls flow". However, this definition results in the "foramen ovale" being classified as a "heart valve", since it undoubtedly is located in the heart and functions as a valve (to switch between the foetal and post-natal patterns of circulation) . Such experiences led GALEN to the rule that, in general, named body parts would be treated as natural kinds and represented as "elementary". Exceptions include cases of generic parts that can be selected, e.g. "lobe of liver" (see 4.1.3) and "named" entities (see 3.1.4).
2. Whether it is useful to define the category, with respect to the needs of the applications expected to be supported within the scope of the model. Some categories are simply not worth the trouble to define, even though definition might be possible. This is particularly true if constructing the definition would necessarily involve the creation and modelling of new categories otherwise very much outside the scope of the ontology and applications. For example, although a sufficient definition of "stroboscope" might be possible in a much broader ontology, within the scope of GALEN it suffices to leave it as elementary.

2.2.2 Explicitness and Orthogonal Taxonomies: "Normalising" the Ontology

Potentially, it should be possible to re-arrange the ontology along any axis. In a description logic, this corresponds to saying that it should be possible to classify any entity according to each of its stated properties. Therefore, all properties must be represented explicitly and independently, even at the cost of apparent redundancy. For example, GALEN maintains that the indications for a drug should be represented separately from its actions even though one can often be inferred from the other, *e.g.* that an indication of "relief of bronchoconstriction" should be represented separately from the action of "bronchodilatation".

GALEN formulates this as the "principle of orthogonal taxonomies" [43, 45], and it has since been elaborated into a general rationale and methodology for "normalising" ontologies [44]. Interestingly, there is a close analogy between the "principle of orthogonal taxonomies" and Smith's advocacy of single inheritance for the "is-a" relation [70], based on entirely different considerations.

2.2.3 Self-standing Entities and Modifiers

The entities in the GALEN ontology can be divided into two kinds:

1. Those that represent things that can exist on their own, *e.g.* physical objects, processes, ideas, etc. Sowa [73] after Pierce terms these "first class objects, whilst Welty and Guarino term them "sortals" [20, 83]. In more recent work the authors of this manuscript have termed them "self-standing entities" [44].
2. Those that only make sense when linked to some other object *e.g.* modifier, modalities, or notions such as "collection of". "Modifiers" are notions such as

"severe", "soft" or "short" that describe other entities and specialise them further. "Modalities" are notions such as "presence", "uncertainty", "family history" etc. that take their meaning from the kernel entity. Sowa [73] after Pierce terms such entities "seconds" and "thirds"

The most important principled differences between self-standing entities and modifiers in GALEN's Common Reference Model are that:

1. Lists of self-standing 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, even in formalisms supporting such inferences.
2. Lists of modifiers may be 'closed', *i.e.* may be assumed to be complete so that inferences of the form "not raised or normal, therefore depressed" can be justified logically, although they must be used with care clinically.

For both technical and clinical reasons, GALEN treats all lists of categories as 'open'. It never makes inferences such as "not absent implies present" on the grounds that this risks imputing a degree of logical rigour to clinicians' statements which is rarely intended. Nonetheless, it maintains the distinction between self-standing entities and modifiers as a top level dichotomy in the model.

2.2.4 Reified Relations[6] or "Features"

The choice of what should be represented as an "Attribute" or "semantic link"[7] is less simple than it seems, since any attribute can be reified (or "nominalised") into a category, *e.g.* in GRAIL notation:

Disease which hasSeverity severe

might also be expressed as

Disease which hasFeature (Severity which hasState severe)

In the second form, the attribute *hasSeverity* has been 'reified' to the category *Severity* plus two subsidiary attributes, *hasFeature* and *hasState.* Such reified attributes, such as *"Severity",* are known in GALEN as *Features*.

Given that this transformation is always possible formally, in the extreme a system could be built with just two attributes (semantic links) for modifiers – *hasFeature* and *hasState*. How, then, should the decision be made as to which attributes to reify? GALEN offers two criteria

1. Need for further description of the attribute – In most formalisms including GRAIL, attributes cannot themselves be described except in predefined ways in the formalism, such as being transitive or having a parent super-attribute in the kind-of hierarchy. Therefore, if the 'fact of being linked' may need to be described, even if only in a few cases, then the attribute representing the link must be reified to a *Feature*.

[6] Note that the word "reify" is used differently with different technical meanings in each of the RDF and Topic Map communities.

[7] Known variously as a "semantic link" (CEN TC251/ISO 215), "property" (OWL), "role" (most other description logics) and slots (frame systems).

2. Consistency of representation – If there are a series of properties that appear analogous, it is almost impossible for authors to maintain a system in which some are represented as an attribute and some as a *Feature*. Therefore, if any must be described as in a) and therefore reified, then all similar attributes should be reified.

In practice, GALEN reifies all modifiers such as severity, height, body temperature, etc. but not 'selectors' such as right in "right hand" about which nothing more can be said. *Features* in GALEN correspond closely to what Welty and Guarino term "qualities" [83], and GALEN's values and *States* to what they call "quale".[8]

2.2.5 Dualities

Many medical concepts come naturally in dualities, and it is not always obvious which should be represented as primary. For example, the "process of ulceration" has as its outcome "ulcer lesions". Should the process be defined in terms of the lesion or *vice versa*? Or should both be treated as elementary and related by necessary but not sufficient conditions? The choice is unclear and possibly arbitrary, but it needs to be made consistently if classification is to work consistently, since "lesions", "processes" and "situations" are different kinds of categories and one will never be inferred to be a subclass of another. GALEN represents the process as elementary and defines the lesion in terms of the process in virtually all cases, even when this requires some linguistic awkwardness (*e.g.* what is the name of the process by which a bullous lesion is formed?).

2.2.6 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, domain independent, categories were seen as making little difference to classification and inference, since most inferences depended more on consistency of expression locally than on top level constraints. Experience has largely confirmed this view technically but, paradoxically, refuted it pragmatically with respect to the development process. An agreed and understandable top level ontology has proved essential to allow groups to co-operate effectively.

However, just as all ontologies are approximations, so all high level ontologies are to some degree arbitrary. There were several candidate starting points early in GALEN's development – PENMAN[3], Cyc [21, 31], traditional schemes from Artificial Intelligence and linguistics such as those deriving from Shank [64] and Sowa [73]. GALEN's top level categories were originally adapted from those in early versions of Cyc [31]. Of recent developments, they are closely related to those in Guarino and Welty's DOLCE [33] and conform to most of the precepts advocated in their OntoClean methodology [19].

In addition, it seems that each major field such as medicine requires one or two very high level abstractions which are broad disjunctions cutting across the traditional boundaries of top level categories. In GALEN, the category *Phenomenon* and the attribute *involves* are designed to range over anything that is, or might become, pathological – in common parlance anything that might be or become a disease, disorder or condition.

---

[8] For a recent discussion of these issues in the context of OWL, see the Semantic Web Best Practice Committee's note on "n-ary relations", http://www.w3.org/TR/swbp-n-aryRelations/.

2.2.7 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 rise to problems when describing congenital malformations and mutilations. There are at least three complementary approaches to this problem:

1. To adjust the interpretation of the attributes and categories. For example, GALEN interprets the *has Division* attribute in such a way that the “*Hand isDivisionOf Arm”* is true even if the hand is severed from the arm. Since we may still wish to represent information about the missing hand relating it to its original owner, this is the best ‘logical approximation’.
2. To model both normal and abnormal, but use the interface and related mechanisms to limit the initial display view only to the normal conformation. The PEN&PAD/Clinergy systems based on GALEN[30, 38] used this approach in many places.
3. To model anatomical normality explicitly, so that almost all statements become statements about “normal hand”, “normal body”, etc. Although elegant, and discussed at greater length in [47, 53], the additional complexity in both modelling and computation combined with the large size of the GALEN ontology made this approach impractical.

Normative statements give more difficulty when applied to procedures and treatments. Consider O’Neil’s classic example, “Insertion of pins in the Femur” [39], which is almost always performed only in order to fixate a broken femur. If a classifier is to infer that it should be classified under “Operations to fixate long bones”, then the information about the goal of the procedure must be added to the description of the method. However, to do so risks imputing unstated intentions to the clinicians using the terminology. GALEN is cautious about adding such unstated normative descriptors, but has found that some cannot be avoided if the classification expected and intended by users is to be maintained.

2.3 Logical Issues

2.3.1 Negation and Uncertainty

Negation and uncertainty lead to difficulties for at least four reasons:

1. The meaning of negation and uncertainty in clinical observations is unclear. For example, where no mention of diabetes exists in a medical record, 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"? Whatever choice is made, can we count on doctors to use it consistently? Dare we therefore support or depend on it?

2. The scope of negation is often unclear. At least three cases must be distinguished: a) "It is not the case that the patient has X"; b) "The patient has non-X" *e.g.* apyrexia (no fever), atonia (no muscle tone), amastia (no breast); and c) "The patient has X but not some specific kinds of X", *e.g.* "idiopathic hypertension" (hypertension but not any of a list of recognised kinds), "Non-toxic goitre", (goitre but not any of the toxic varieties") or "non-A non-B hepatitis" (hepatitis but not that caused by either the hepatitis A or B virus).
3. Adding negation and uncertainty to formalisms increases their computational complexity and makes canonization difficult. Even ontologies based on underlying formalisms that support negation may choose not to use it.
4. Negation and uncertainty are often represented in information systems models *e.g.* the HL7 Reference Information Model (RIM)[9]. If negation can be represented both in the information system and in the ontology, then the meaning of all possible combinations of negations in the two systems must be defined. (See[50, 54, 55].)

GALEN's GRAIL formalism does not support negation, but the GALEN Common Reference Model includes constructs such as "presence" and "absence" which provide a limited 'work around' and that can be qualified by an uncertainty.

2.3.2 Defaults and Indexing

The definition of "B is a kind of A" in formal logical representations is that "All Bs are As". Hence, all of the properties in the definition and description of 'As' must also apply to 'Bs' without exception. Adding exceptions to such logical patterns has had little success [15], although it remains an area of ongoing research. This contrasts with most, although not all, frame systems in which default values for a 'slot' (equivalent to a GRAIL attribute) can be both inherited and overridden.

However, if additional facts are indexed by an ontology that conforms to this logical definition of 'is kind of', then it is still possible to use the ontology in conjunction with other inference mechanisms to reason about defaults and exceptions. For example, a logical subsumption hierarchy from an ontology of drug classes can be used to index potential side effects, even though some side effects are subject to exceptions [71] The scaffold provided by the subsumption hierarchy can be used to select the most specific candidate side effects using the standard "Touretzky distance measure" [78].

GALEN refers to such indexing statements as "extrinsic" because they do not affect the classification and are therefore not part of the ontology proper but rather use the ontology as, in Wood's [85] phrase, a "conceptual coat rack" on which to hang other information.

GALEN's experience is that if the taxonomies are properly orthogonal – *i.e.* if the ontology is normalised – the set of candidate values usually has exactly one member.

---

[9] http://www.hl7.org.

If it does have more than one member, then GALEN treats this as a signal that other reasoning methods and knowledge are required.

2.3.3 Definitions and General Inclusion Axioms

Unlike most DLs of its generation including that used in SNOMED-CT, GRAIL allows defined categories ("classes") to be further described by "necessary statements". This means that GALEN's authors do not have to choose between making all of the characteristics of an entity part of its definition (*i.e.* necessary *and* sufficient) or all merely necessary. For example, consider the notion that "severing of an artery" causes "haemorrhage". One would not want "causing haemorrhage" as part of the definition of the severing of an artery – *e.g. Severing which actsOn Artery* – because then we should have to state explicitly that "severing the aorta" had caused a haemorrhage before a machine could classify it as a "severing of an artery". On the other hand, we would want the ontology to include the information that all such injuries are kinds of injuries that cause haemorrhage. Such additional necessary but not sufficient conditions are known in description logic as "general inclusion axioms".

2.3.4 Embedded Expressions

If a category's representation depends on its use, then this limits its re-use. Categories 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, substances to be injected or drained, or specimens in a pathology examination. Since many of these categories are themselves composite, a primary requirement on the GRAIL language was that it allow definitions to be recursively embedded within other definitions to any degree required. For example, GRAIL supports expressions such as "upper part *of* third segment *of* middle lobe *of* right lung". Such embedding is impossible in most frame languages and has not been used in SNOMED-CT, beyond the mechanism of "role grouping" for a single level of embedding.

2.3.5 Transitive Attributes and Inheritance Across Transitive Attributes

Part-whole relations, causal links, and connections are all transitive. Some other attributes, though not themselves transitive, are 'inherited' across these transitive attributes. Establishing the pattern of transitive relations and the inheritance along them is a key part of any ontology of medicine [46].

GALEN's original primary use for transitive attributes was for part-whole relations; its original use case of inheritance across transitive attributes was for representing the patterns "The disease of the part is a disease of the whole" and "The procedure on the part is a procedure on the whole". These two specific cases might now be implemented instead by SEP triples [24, 25] or one of their variants [42]. However, GALEN also uses inheritance across transitive attributes to support several other clinically important inferences in an otherwise relatively 'weak' description logic. For example:

1. 1)In the representation of syndromes, to represent the fact that the presence of a syndrome implies the presence of each disease in the syndrome.

2. 2)In the representation of procedures, to represent that a global procedure acts on all of the structures acted on by its subprocedures.
3. In the representation of anatomy, to represent that where a subbranch of a larger vessel supplies blood to a particular structure, then this implies that the larger vessel also supplies blood to that structure

In GALEN, such axioms are implemented by the use of the *specialisedBy* construct, equivalent to "right identities" in SNOMED-CT's representation.

In addition, GRAIL supports a construct for 'single valued' transitive attributes, which is interpreted as indicating that the transitive attribute must form a tree. This avoids the need to provide non-transitive "direct" subattributes of transitive attributes.[10]

2.3.6 Issues Minimally or Poorly Represented

1. *Adjacency and spatial/temporal reasoning.* A "fracture of the tibia and fibula" makes sense; a "fracture of the tibia and humerus" does not. GALEN provides very limited support for this type of reasoning, although there have been experiments with several work arounds. Likewise for more complex relations involving spatiotemporal reasoning and its interaction with plausible mechanisms of injury or pathophysiology. It is assumed that these will be dealt with either by the information model or by separate reasoners outside the central terminology/ontology.

GRAIL	OWL	DL	Paraphrase
(C which prop C_1) name CN-	CN class C complete restriction (*prop*_ someValuesFrom C_1)	$CN \doteq C \sqcap \exists prop.C_1$	CNs are defined as any C which *prop* some C_1
C topicNecessarily prop C_1	subclassOf(C, restriction(*prop*_ someValuesFrom C_1))	$C \sqsubseteq \exists prop.C_1$	All Cs prop some C_1
<*$prop_1$ C_1 $prop_2$ C_2 ... $prop_n C_n$*>	intersectionOf(restriction(*prop* $_1$ someValuesFrom C_1), restriction(*prop* $_2$ someValuesFrom C_2) … restriction(*prop* $_n$ someValuesFrom C_n))	$\exists prop_1.C_1 \sqcap \exists prop_2.C_2 \sqcap \ldots \sqcap \exists prop_n.C_n$	*prop* $_1$ some C_1 and *prop* $_2$ some C_2 and …. and *prop* $_n$ some C_n
$prop_1$ specialisedBy $prop_2$; $prop_1$ refinedBy inverse $prop_2$	--	$prop_1 \circ prop_2 \sqsubseteq prop_1$ equivalent to $prop_2^{-1} \circ prop_1^{-1} \sqsubseteq prop_1^{-1}$	Any C_1 that *prop* $_1$ some C_2 that *prop* $_2$ some C_3= also *prop* $_1$ some C_3, or equivalently: Any C_1 that inverse *prop* $_2$ some C_2 that inverse *prop* $_1$ some C_3 also inverse *prop* $_1$ some C_3.

Fig. 1. Grail modeling **constructs**

[10] See Simple Part-Whole Relations in OWL ontologies, Rector & Welty (eds.) http://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/

GRAIL	OWL	DLs (FaCT/Racer)	OntoClean/ Dolce	Logic
Category	Class	Class	(unary) Predicate	unary predicate / Type
Attribute	Property	Role	Relation	Binary predicate/ Relation
necessary statement *topicNecessarily*	subclassOf() axiom	"General inclusion axiom" *implies* (⊑)	→	→

Fig. 2. Comparsion of Grail and other vocabularies

2. *Numerical conversions, calculations and other 'non-terminological' reasoning.* There are numerous services that users might naturally expect to be packaged with a terminology but which require entirely different types of reasoning from logical classification based on definitions, descriptions and first order logic. The most obvious of these are conversion between different unit and coordinate systems. GALEN's intention has always been to package these services separately within the 'terminology server', and the architecture provides for them although, in practice, none have been implemented. However, they are strictly excluded from the "ontology" or Common Reference Model (CRM).

3 The GALEN Upper Domain Ontology

The GALEN Common Reference Model is presented here using the notation of GALEN's GRAIL language. However, the presentation is intended to be sufficiently general to allow comparison and potential harmonisation with other clinical ontologies such as that of SNOMED-CT[11], the Digital Anatomist Project [34, 63], where appropriate with the Gene Ontology and other ontologies from the Open Biomedical Ontologies (OBO) group [76, 77][12], with more language oriented work such as that of Zweigenbaum [91] or Hahn [23], or with more general upper ontologies such as DOLCE/OntoClean from Guarino and Welty [33, 74, 83], SUMO[13] or Smith's Basic Formal Ontology and its biological adaptations [13, 70]. The full ontology is available from the *Open*GALEN web site, http://www.opengalen.org, and a detailed description of the GRAIL language is available in [46]. A short summary of GRAIL notation as used in this paper and its equivalents in OWL and standard German DL notation along with notes on unusual features is given in Figure 1, and additional vocabulary comparisons are given in Figure 2. The GALEN vocabulary is explained in the text.

This paper focuses on the issues raised and is not intended as a guide to the current implementation. In some cases, the constructs and language used reflect more recent developments not fully implemented in the existing resources available from *Open*GALEN. Where there are significant departures from the actual implementation, they are noted in the text.

[11] http://www.snomed.org
[12] http://http://obo.sourceforge.net/
[13] http://suo.ieee.org/

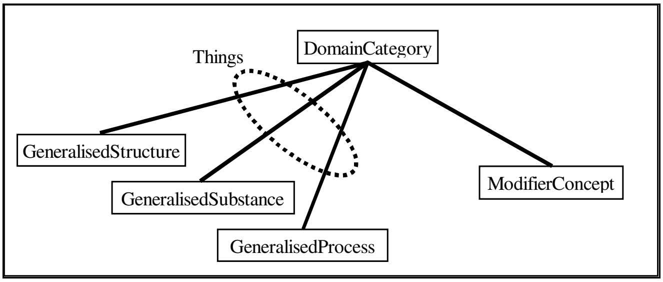

Fig. 3. Primary structure of Galen's Toplevel Categories

The ontological patterns described in this paper are for the raw, underlying ontology. GALEN treats this ontology as an 'assembly language' that few users ever see. The goal of this underlying ontology is to be unambiguous and result in correct classification. Intuitive presentations to users are dealt with via intermediate representations and tools [48, 56, 60] which are outside the scope of this paper.

3.1 The Top Level Categories

3.1.1 Top Level Distinctions

The primary structure of GALEN's top level categories is shown in Figure 3. GALEN's top level distinction is between self-standing entities, or *Things*, and everything else, termed *Modifier Categories. Things* are roughly equivalent to 'sortals' in DOLCE and are further divided into

GeneralisedStructure	abstract or physical discrete *Things* with parts that exist at particular times, e.g. bodies, organs, cells,…
GeneralisedSubstance	abstract or physical continuous *Things* with parts which exist at particular times, e.g. tissues, fluids,…
GeneralisedProcess	changes which occur over time, e.g. metabolic processes, procedures, …

These distinctions are now common currency although under different names. *GeneralisedStructure* and *GeneralisedProcess* together are approximately equivalent to "endurants" in DOLCE , or "continuants" in the BFO and many other ontologies. *GeneralisedProcess* is equivalent to "occurents" in the BFO and "perdurants" in DOLCE. *GeneralisedSubstance* corresponds to "Amount of matter" in DOLCE but has no equivalent in BFO. The structure was originally adapted from Lenat and Guha [31], but where they maintain a distinction for processes analogous to that between *GeneralisedStructure* and *GeneralisedSubstance – e.g.* between "the digestion of a meal" and "the activity of digestion"– GALEN does not, because knowledge engineers and users found it to be confusing and difficult to maintain reliably. Neither DOLCE nor BFO support this distinction nor, it appears, does the current version of OpenCYC.[14] For different reasons, the notion of "*Thing"* as the common parent of *GeneralisedStructure* and *GeneralisedProcess* was left implicit, as its labelling led to arguments about language. GALEN does not make the distinction between "function"

[14] http://www.cyc.com/doc/

and "process", i.e. between the potential for a process to occur and an occurrence of the process, as made in BFO and DOLCE.

3.1.2 Modifiers

The first level break down of *ModifierCategory* falls into:

- **Aspect and Modality**[15]
  - *Aspect* — 'modifiers proper' that refine a category, e.g. size, shape, age, laterality, etc[16].
  - *Modality* — Separate notions that take part of their meaning from the primary things, e.g. family history of, risk of, history of, etc.
- **Other categories that are dependent on self-standing entities for their full meaning**
  - *Role* — sometimes arbitrary categories used to make elementary taxonomies orthogonal, e.g. *DoctorRole*, HormoneRole, *DrugRole*, etc.
  - *Collection* — set, system, multiple, etc. GALEN's collections are not mathematical sets but rather various forms of general collection such as vertebrae, the cells in the liver, etc. GRAIL supports no special operations on collections.
- **Miscellaneous categories with special significance or behaviour**
  - *Unit* — units of measure, *e.g.* mg, ml, day, …

Of the above, the most complex is *Aspect,* which is further subdivided into:

- *Feature* — reified attributes (see 0) representing mutable properties *e.g.* severity, duration, etc. To have meaning, *Features* must be further refined either by one or more *State*s in a "Feature-state pair" (*e.g.* T*emperature* *which* *hasState hot*) or by the entity that it is a property of (e.g. *Length* *which* *isLengthOf Bone*).
- *State* — (usually) closed sets of qualitative 'value's that may be assigned to *Feature*s, *e.g. mild, moderate, severe*.
- *Quantity* — used to refine *Features* with quantitative values, including numerical magnitudes and units or levels
- *Selector* — immutable properties *e.g.* laterality (left/right) and position (upper/middle/lower) etc. of anatomical parts. Selectors identify a specific entity rather than modifying it[17].
- *Status* — Modifiers other than selectors that are not reified; many are used to support special inference in the model or in applications using the model, *e.g. normal/nonNormal, countable/indefinitely Divisible/mass,* and various topological indicators[18].

---

[15] The labels "Aspect" and "Modality" were arrived at after much internal discussion. "Modality" corresponds roughly to what SNOMED-CT refers to as "Axis modifying qualifiers" and "Aspect" to "Non-axis modifying qualifiers".

[16] Corresponds approximately in SNOMED-CT to "non-axis changing qualifiers".

[17] In terms of OntoClean, selectors are part of what gives an individual an identity. A left hand cannot cease to be of laterality left without becoming something different.

[18] Also used as a ragbag for qualitative values not currently represented as feature-state pairs.

Most mutable properties except *Status*es are reified in GALEN to *feature-state-*pairs, *e.g. Disease which hasFeature (Severity which hasState severe).* By contrast, *Selector*s are immutable and always linked directly to the entity they modify by a single attribute, *e.g. Hand which hasLeftRightSelector rightSelection. Status* in GALEN is defined by engineering rather than ontological principles; it includes primarily immutable properties such as an organs topology but also the sometimes mutable property of whether or not a given entity is *nonNormal* and/or *pathological*.

The special Quantity[19] *Level* is used amongst other things to represent the recurrent pattern in departures from expected values first pointed out by Shahar [68]. *Level* takes a series of subattributes of *hasState – hasAbsoluteState, hasChangeInState, hasTrendInState, hasRelativeLevelState,* and *hasExpectedLevelState.* This allows the expression of complex notions such as "temperature with an absolute state of 39°C, which is falling, but which is still elevated (*i.e.* higher than expected)".

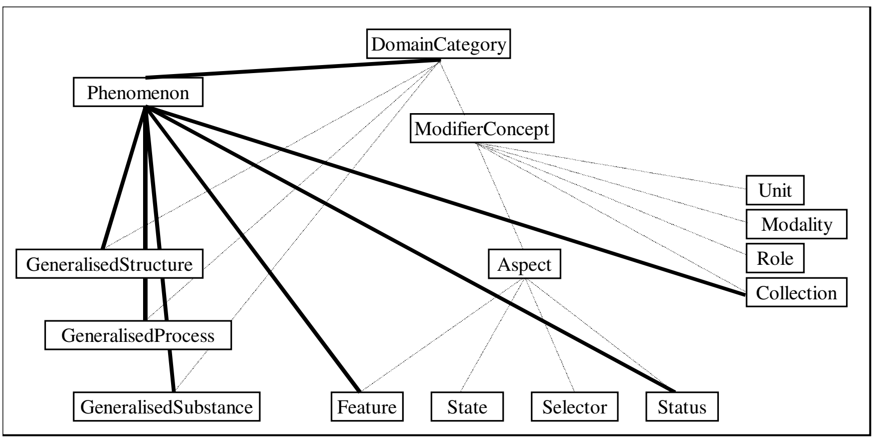

Fig. 4. Sescondary structure of Galen's top level categories

3.1.3 *Phenomenon* – Secondary Structure for Top Level Categories

As with many ontologies oriented to a particular domain, GALEN requires a very high level disjunctive category to allow representation of key clinical generalisations. In GALEN this category is labelled *Phenomenon*, the common ancestor of anything that can be, or can be modified to be, worth noting clinically as either *nonNormal* or *pathological.* GALEN lacks an operator for disjunction, so Phenomenon is added manually, as shown in Figure 4, as the common parent of the subsidiary categories. As defined, this is clearly too inclusive to meet GALEN's original goal of representing all and *only* what is clinically sensible. However, the effort to tighten the constraints whilst avoiding arguments over issues such as whether or not an infected prosthesis can be *pathological* has not so far been warranted.

[19] Whether "*Level*" should be a *Quantity* or a *Feature* has been a matter of some controversy but is without obvious consequences for the inferences to be made.

3.1.4 Breaking Up Long Lists: The NAMED… Convention

The principle of orthogonal taxonomies combined with the principle that all anatomical entities be treated as natural kinds, results in a broad flat hierarchy of elementary categories that is difficult to work with. For convenience, GALEN breaks this hierarchy up by introducing categories such as *NAMEDArtery*, *NAMEDJoint*, NAMEDSensoryPart etc.

3.2 Top Level Attributes[20]

3.2.1 Primary Distinctions

In GRAIL, as in many but not all other description logics, one attribute ("role", "property") can be a kind of one or more others, just as one category can be a kind of another. *DomainAttribute* is the root of the attribute polyhierarchy, and it breaks down into three major branches, each of which will be discussed in turn. (Note that all attributes in GRAIL have inverses that have been omitted in this paper for clarity. By convention, attributes and their inverses are named by analogy to *isPartOf* and *hasPart*.)

ConstructiveAttribute	Relations between self-standing entities (Things), ie. GeneralisedStructures, GeneralisedSubstances, and GeneralisedProcesses
ModifierAttribute	Relations between Things and ModifierCategories
TemporalAttribute	Relations between Statuses involving time (deliberately weak, see 2.3.6)
WrapperAttribute	Used in ClinicalSituations (see 5)

3.2.2 ConstructiveAttribute

ConstructiveAttribute further breaks down into three primary subattributes plus the domain specific secondary attribute, *LocativeAttribute (*alias *involves).*

PartitiveAttribute	Part-whole relations –e.g. *isDivisionOf* - see **4.1.1**
StructuralAttribute	Non-partitive relations e.g. *isServedBy*, *isBranchOf*, *isSpaceDefinedBy*, etc.
FunctionalAttribute	Functional relations such as *isFunctionOf*, *actsOn* etc.
LocativeAttribute (involves)	A heterogeneous disjunction of locative, purposive, functional and causal relations, e.g. *isConsequenceOf*, *isFeatureOf*
- *hasLocation*	The relation between a disease and the anatomical or physiological entity in which it is localised. NB does not imply physical location.[21]

---

[20] "Properties" in OWL; "Roles" in standard DLs; "relations" in standard mathematical terms; "semantic link types" in CEN TC251/ISO TC215.

[21] Because the naming of the attribute "hasLocation" has led to confusion in alignment with other ontologies, for conversions and other external uses it has been renamed to "hasLocus". Approximately equivalent to the SNOMED-CT "site".

The key construction in most medical entities is to localise a disease or procedure to an anatomical or functional entity or to one of its parts. Correspondingly, the most common pattern in GALEN for disease or procedure entities is:

Disease/Procedure which LocativeAttribute BodyStructure/Process

specialisedBy		**Example**
hasLocation	*isDivisionOf*	*Disease hasLocation (Part which isDivisionOf Whole) →* *Disease hasLocation Whole*
isComponentOf	*isSubdivisionOf*	*Bone isComponent of (Finger which isDivisionOf RightHand) →* *Bone isComponentOf RightHand*
isLayerOf	*isSubdivisionOf*	*Skin isLayerOf (Hand which isSubdivisionOf UpperExtremity) →* *Skin isLayer of UpperExtremity.*
isBranchOf	*isLinearDivision Of*	*CoronaryArtery isBranchOf (AscendingAorta isLinearDivisionOf ThoracicAorta) →* *Vessel isBranchOf ThoracicAorta*
serves	*isDivisionOf*	*BloodVessel serves (Part which isDivisionOf Whole) →* *BloodVessel serves Whole*
contains	*isLocationOf*	*Abdomen contains (Liver isLocationOf Tumour) →* *Abdomen contains Tumour*
actsOn	*isFunctionOf*	*Drug actsOn (PathologicalProcess isFunctionOf Organ) →* *Drug actsOn Organ*
actsOn	*makesUp*	*Process actsOn (Tissue makesUp Liver) →* *Process actsOn Liver*
contains	*isLocationOf*	*BodySpace contains (Organ isLocationOf Lesion) →* *BodySpace contains Lesion*

Fig. 5. Important uses of the specialisedBy construct indicating inheritance along a transitive role and equivalent to SNOMED-CT right identities

3.2.3 ModifierAttribute

The modifier attributes and modifier categories are intimately tied, one main branch of the attribute hierarchy for each branch of the *ModifierCategory* hierarchy: *modalityAttribute, RoleDesignatingAttribute, CollectionAttribute, UnitAttribute* and the attributes related to *Aspect – isFeatureOf, isStateOf, QuantityAttribute, SelectorAttribute,* and *StatusAttribute.*

Two limitations of GRAIL lead to a proliferation of subattributes that are of no ontological significance but can obscure the overall structure.

1. Cardinality can be controlled only at the level of attributes – in modern parlance “qualified cardinality restrictions” are not supported. GRAIL shares this feature with OWL (in all its flavours) and SNOMED-CT, although not with most modern DLs. Therefore, separate subattributes must be used for single valued variants of attributes.
2. The GRAIL category hierarchy represents most modifiers by reifying the relation to a kind of *Feature*. Since each individual can have many *Features*, but only one of each kind of *Feature*, a separate subattribute of *hasFeature* is required for each *Feature – hasTemperatureFeature*, *hasHeightFeature*, etc.

3.2.4 Structure of Inheritance Across Transitive Attributes

In addition to the attribute hierarchy as described above, GALEN provides the *specialisedBy* construct as described in 0 for inheritance of attributes across transitive roles. Some of the most important *specialisedBy* axioms are given in Figure 5.

3.2.5 Additional Uses of the Attribute Hierarchy

Two further uses of the attribute hierarchy deserve special mention. The first two are logical; the third is ontological.

1. To allow single-valued and multi-valued variants of an attribute. Logically, the single-valued variant must be a descendent of the multi-valued variant, and its purpose is signalled by the infix "specific" or "specifically", *e.g.* *hasSpecificConsequence* or *actsOnSpecifically.* Such "specific" attributes are often used to indicate a main, or primary action, cause, etc.
2. As a workaround for the lack of 'shared variables' in GRAIL (as in other description logics). GRAIL provides no mechanism to represent 'X *containedIn* Y ← X *part of* Y'. GALEN achieves an approximation to this inference by the attribute *isPartitivelyContainedIn*, which a descendant of both *IsDivisionOf* and *Contains*.
3. To allow very general queries, such as "disorders of the heart". *LocativeAttribute* (also known as *involves*) has been steadily generalised in the course of the project until it has become the analogue of *phenomenon,* a domain specific disjunction of the attributes needed for high level generalisations and queries. It is worth noting that, in this very general form, *LocativeAttribute* subsumes causal relations since, for example, classifying "spider angiomata" under "phenomena involving liver disorder" is appropriate. Similarly, rheumatic heart disease *involves* bacterial disease as well as a heart disease since the lesions located in heart are in response to an infection caused by bacterium.

4 The GALEN Common Reference Model

4.1 Anatomy

One of the key aspects of any biomedical ontology is its representation of anatomy. Because GALEN has been used most extensively for developing terminologies of surgical procedures, its anatomy representation is considered the best developed and tested and is presented in detail below.

4.1.1 Physical Part Whole Relations and Physical Connection

There has been much study of parts and wholes – in GALEN's parlance "partitive relations"– in AI generally, *e.g.* [40, 84], and in description logics more specifically [1, 2, 41]. An entire subfield of philosophy and linguistics - "mereology" - is devoted to their study [4, 9, 82]. Technical details of how GALEN's mechanism for inheritance across transitive properties is applied to parts and wholes, and how this relates to other formalisms, can be found in [42]. Since GALEN's ontology was established, variants on Schulz and Hahn's SEP triples formalism have been widely used as a means to implement related ideas [24, 25, 65-67]; these will be further considered in the discussion section.

As anatomy is physical, we deal here only with partonomy as it relates to physical things[22]. The basic axioms of the GALEN model of partonomy are as follows:

Rule 1) All primary partitive attributes between discrete objects are transitive. This includes isLayerOf on the grounds that anatomical layers are always concentric [56].

Rule 2) Diseases/disorders/procedures of/on a part pertain also to the whole

Rule 3) "Connection" is transitive23 but not always partitive. A combined attribute, isPartitiveConnectionOf, is provided for cases where it is partitive;

Rule 4) "Branching" is neither partitive nor transitive, although because isBranchOf is refined along isLinearDivisionOf (See

Rule 5) and Section 2.3.5 above), branches of linear divisions are branches of the whole, e.g. branches of the infrarenal aorta are classified under branches of the abdominal aorta.

Rule 6) Connected physical sets such as the "digestive tract" are distinct from functional systems such as "the digestive system"

Rule 7) Membership in collections is not partitive, contrary to [40, 84].

GALEN then classifies the range of possible part-whole relationships between discrete physical parts along several axes, with strong constraints based on the topology of the arguments and whether they are *Structures* (discrete) or *SubstanceOrTissues* (continuous/mass).

isDivisionOf	The most general partitive attribute
- *isLinearDivisionOf*	Relates any two topologically linear structures, *e.g.* between an arterial segment and the artery
- *isSurfaceRegionOf*	Relates a two-dimensional structure to a three-dimensional structure, such as between an organ and its surface
- *isSolidRegionOf*	Most general relationship between any two three-dimensional structures.
- - *isLayerOf*	Relates things like skin or muscle or periosteum that occur in all divisions of an entity to that entity.
- - *isSolidDivisionOf*	Relates all other three-dimensional entities, ie wherever the relationship is not 'layer-of'
- *isComponentOf*	Relates discrete things like joints, ligaments and organs that occur only in one or more divisions of an object
- - *isFunctionalComponentOf*	Participates in a specialisedBy axiom such that functions of the part are also functions of the whole.

---

[22] Physical endurants/continuants in DOLCE/BFO's parlance.

[23] Other authors take connection as only symmetric, and not transitive. GRAIL does not support symmetric relations, while GALEN's "connection" corresponds to the transitive closure of all direct and indirect connectivity, where a true 'directly connected to" relation would indeed not be transitive.

These partitive attributes are further related by the following rules:

Rule 8) Components of any discrete part are components of the whole, *e.g.* the chordae *of* the leaflets *of* the valves *of* the ventricles are components of the heart.

Rule 9) Layers of divisions are layers of the whole, *e.g.* the skin of the hand is a kind of skin of the upper extremity.

Rule 8) above is a pragmatic approximation and the one case in GALEN where part-hood and subsumption are deliberately conflated. The rule should be: "Layers of divisions *are divisions of* the corresponding layer of the whole", *e.g.* "The hand is a division of the upper extremity; therefore the skin of the hand is a division of the skin of the upper extremity." Unfortunately, this rule is outside the expressivity of description logics[24] [42]. In practice, we have not discovered any errors due to this subsumption at the gross level of anatomy needed for GALEN's focus on diseases and procedures, although it would not be adequate for some parts of developmental anatomy.

Rules 2,4,7 & 8 are implemented by the use of the *specialisedBy*[25] construct for propagation along transitive roles (see 2.3.5).

One rule was not properly implemented in GALEN although it appears in various places in the documentation, because the distinction between discrete components and subdivisions was not fully implemented.

Rule 10) Layers of discrete components should not be layers of the whole (e.g. the cartilage layer of the tibial plateau should not be a kind of layer of the knee joint)

One further rule would be required in most other formalisms that - unlike GRAIL - do not support restrictions of transitive attributes *e.g.* to strict trees.

Rule 11) All transitive attributes have a direct non-transitive subproperty.

4.1.2 Regions

The problem of describing what clinicians refer to as regions of the body poses significant headaches for a logic based ontology, not least because regions have borders that are either ill defined or defined differently by different experts and even different text books. In addition to these difficulties, the following challenges were encountered:

1. Regions named identically with the primary structure that they contain, *e.g.* 'knee' may refer either to the knee joint or the knee region. GALEN treats both "regions" and associated primary structures as primitives, with the structure being necessarily *isStructuralComponentOf* the region. (Note that GALEN's naming convention assigns the 'simple' name to the surface region, *e.g.* "Chest" or "Knee", whereas the FMA assigns it to the associated structure. GALEN's *Knee* corresponds to FMA's "Region of the knee"; GALEN's *KneeJoint* is FMA's "Knee")
2. Regions defined as those areas of (unspecified) tissue that have a particular, though often loosely bounded, spatial relationship to some named structure (*e.g.* paracolic

[24] It requires at least three variables to express the rule in formal logic. It is therefore is outside F2, first order logic with two variables. All DLs are subsets of F2.

[25] Often expressed in the actual source files by its converse, refinedAlong – see Fig 1.

gutter) or are simply 'near' them (*e.g.* perianal abscess). GALEN defines such structures using the special attributes *hasProximity* (*e.g.* perianus), *isParallelTo* (*e.g.* paramedian line), *isColinearWith* (transurethral route) and *passesThrough (e.g.* percutaneous route*).*

3. Regions named according to their clinical significance and whose boundaries cannot be inferred on the basis of purely anatomical relations: *e.g.* the "precordium" is the region of the chest specifically associated with observation and auscultation of the heart. GALEN represents such structures as primitives, though these may be further described using one or more of the partitive, spatial and proximity attributes.

4.1.3 Generic Bits and Pieces

Notions such as "capsule", "spine", or "edge" are widely used in anatomy to identify elements of anatomical structure – *e.g.* "capsule of kidney", "spine of 5th lumbar vertebra", "edge of liver" etc. In modelling such generic notions there are two choices:

1. To represent the generic notions as elementary and the real anatomic structures as defined compositions, *e.g.* "*Angle which isSubdivisionOf Mandible", "Pole which isDivisionOf Kidney"*, etc.
2. To represent each occurrence of the substructure individually as elementary, *e.g. AngleOfMandble, PoleOfKidney,* etc.

In general, GALEN has chosen 1) because a) there seems to be sufficient commonality in notions such as "lobe" or "pole" that some are used for classification, *e.g.* "Lobulated organ" *e.g.* in the FMA, and b) the partitive relationship between such substructures (*e.g.* renal pole) and the anatomical entities of which they are part (e.g. kidney) appears to be defining in nature, rather than only incidentally true[26].

4.1.4 Tissues, Cells and Substances: *Mass*, *Discrete*, and *IndefinitelyDivisible*[27]

Most western languages make a distinction between a) "mass nouns" and "count nouns". Mass nouns such as "water" and "sand" are normally used in the singular; count nouns may be either singular or plural. Lenat and Guha make a corresponding semantic distinction between mass "stuff" and discrete "things" [31]. DOLCE makes the corresponding distinction between "Amount of matter" and "Physical object"; the realist stance of the BFO[28] [4] does not support this distinction.

In GALEN, structures and substances have a countability that is one of:

discrete	Bones, organs, membranes, etc.("countable")
mass	Substances and tissues
- *indefinitelyDivisible*	Cells, grains of sand, etc.
- *indefinitelyMultiple*	(present but not used in existing model)

---

[26] In terms of other philosophical constructs, the notion of "renal pole" can be considered as "analytic".

[27] Actually termed "infinitelyDivisible" etc. in the implemented version.

[28] http://ontology.buffalo.edu/bfo/BFO.htm

The *indefinitelyDivisible* category covers things like cells that are usually treated en masse, as in their count-concentration in a body fluid, but which can have discrete parts. A general mechanism for dealing with granularity has been developed from the GALEN experience, though the issue was never extensively explored in GALEN itself.

4.1.5 Topologies, Cavities, Spaces, Lines and Anatomical Landmarks

All solid structures in GALEN have a topology that may be *topologicalyHollow* or *topologicalySolid.* Being solid is simple; GALEN recognises the following kinds of being hollow:

surfaceHollow	Surface regions such as the "abdomen" which overlie a cavity and are often seen as having things in them
trulyHollow	Properly hollow structures,
- *actuallyHollow*	Not bilayered
- - *closedHollow*	No openings
- - *tubularHollow*	One or two openings. The cavity defined is a *Lumen*.
- *bilayered*	Membranes such as the pericardium and pleura, where the layers are normally in apposition such that the space between them is abolished for all clinical purposes (a potential space)

TrulyHollow body structures define a *Cavity,* which is related to the object that defines it by the attribute *definesSpace,* which is not partitive in the current implementation. The more general notion of a *Space* may be defined or only partly described using the attribute *boundsSpace* to refer one or more objects that are coterminus with any part of the boundary of the space e.g the dura mater and the subarachnoid membrane *boundsSpace* the subdural space.

Clinical anatomy also recognises a large number of points, lines and surfaces. These may be related to other anatomical structures (*e.g.* the pectineal line is the attachment of the pectineus muscle on the femur), while others such as the McBurney's point, the midclavicular line, inguinal triangle and parasagittal planes are treated as structures by fiat. Surgical procedures may reference routes of approach (*e.g.* transoesophageal and percutaneous) that are conceptually linear in nature, though not strictly one dimensional. Furthermore, other notions such as the quadrants of the abdomen have uncertain dimensionality: though they may be defined as planar sections of a planar structure (*e.g.* the anterior abdominal wall) they may also be spoken of as either containing or having as part those structures lying directly below them. Similarly, tubular body structures (however highly convoluted in space they may be) are often referred to as having linear properties – they can have segments.

Therefore, all *PhysicalStructures* are assigned (or inherit) a *Topology*[29] value: *linear*, *laminar* or *solid.* In addition, to deal with cases such as the intestine and quadrants of the abdomen they may be given an *AnalogousTopology*[30] value. The *Topology* governs constraints such as that only a *SolidStructure* may *contain* another

[29] Actually "*Shape*" in *Open*GALEN for historical reasons
[30] Actually "*AnalogousShapeValue*" in *Open*GALEN for historical reasons

PhysicalStructure, and that a *LinearStructure* can only have another *LinearStructure* or a *Point* as a subpart. The *AnalogousTopology* governs constraints such as whether a topologically hollow structure is elongated to be *Tubular* and can therefore have linear divisions.

GALEN recognises two further generic anatomical notions: *SurfaceVisibility* – whether a structure is internal or external – and *PairedOrUnpaired* – whether a structure comes in paired variants (left/right, medial/lateral etc.) and if so, whether they are mirror images of each other (*e.g.* hands) or not (*e.g.* cardiac ventricles).

Finally, whilst GALEN has avoided many of the difficulties inherent in representing non-normative anatomy such as arises through disease (see 2.2.7), even 'normative' human anatomy is inherently sexually dimorphic. GALEN's approach to sexual dimorphism is as follows: all primitive anatomical structures that are specific to one sex only (*e.g.* uterus, testis) are assigned a male or female phenotype value. Structures present in both sexes and with no sexual dimorphism have no phenotype value. Structures with dimorphic variant subforms (*e.g.* breast) carry no phenotype value, but their male- and female-specific variant subforms are instead defined (e.g. *Breast <u>which</u> hasPhenotype male*). Part-whole relations are asserted so that e.g. the sex unspecific *PelvicCavity* is asserted to contain the *Rectum*, but only the *FemalePelvicCavity* contains the *Uterus (*and also, by inheritance from its ancestor *PelvicCavity*, the *Rectum*).

4.1.6 Arbitrary Portions

Clinical descriptions of practical interactions with real anatomy (as opposed to descriptions purely of idealised canonical anatomy) often involve the notion of an arbitrary portion of a named anatomical structure. For example: removal of a *segment* of artery; excision of a *piece* of liver; tumour in the distal third of the humerus. The particular term chosen to denote the portion – *e.g.* segment, chunk or slice – may imply a particular topology of both the target structure and the referenced portion, as well as a particular partitive relationship holding between them.

Building on its strong typing of topology and partonomy as already described, GALEN represents arbitrary portions by means of a single primitive entity: *SolidRegion.* Individual arbitrary portions may then be described as a *SolidRegion* that has a particular partitive relationship with some structure. The topological properties of the portion itself may then be inferred from the topology of the structure of which it is part, and the nature of the partitive relation. Thus, a *Segment* can be defined as a *SolidRegion <u>which</u> isLinearDivisionOf LinearStructure* and must itself have *LinearAnalagousShape.*

4.1.7 Reciprocal Expressions

Unlike most representations of anatomy based on description logic, GALEN contains both statements of the form *B is_part_of A*, equivalent to "All Bs are part of some A" and *A has_part B*, equivalent to "All As have part some B". For example, both "*(All) Hand isDivisionOf (some) Arm*" and "*(All) Arm hasDivision (some)Hand*" can be represented. Such statements are terms "reciprocals". Neither separately implies the other even though *is_part_of* and *has_part* are mutual inverses. Modern "Tableaux reasoners" such as FaCT [27, 28] and Racer [22] are intrinsically exponentially explosive in the face of even small numbers of reciprocal statements. This does not occur in GALEN because the structural algorithms used by GALEN's GRAIL classifiers while incomplete, are efficient even for very highly connected ontologies containing many reciprocals.

This allows GALEN to be much more precise about normative anatomy than systems, such as SNOMED-CT, which confine themselves to "*isPartOf*". However, strictly speaking, it is not true to say that all arms have hands as parts, but only that normatively arms have hands as parts. However, the advantages of being able to express both sides of such relationships outweigh the disadvantages. For purposes of expressing clinical information, the normative interpretation is almost always appropriate, provided notions such as "missing" supplement it.

4.2 Processes and Functions

GALEN uses a relatively simple model of processes and functions. No distinction is made between mass and discrete processes or between processes and events. There are a few primary attributes linking the structure together

actsOn	Processes act on other phenomena: processes, structures, or substances.
hasConsequence	The primary causal attribute – see 4.3.2 below
– *hasUniqueAssociatedProcess*	Links processes to their outcomes. Used in process-outcome duals such as *UlcerProcess* and *UlcerLesion* – see 2.2.5 above
isFunctionOf	Links processes to their actor or the organs or organ system which carry them out
isSubprocessOf	The single primary partitive attribute for processes.
hasGoal	Links processes to their intention (either another process, or a state or a structure)

All of the above functions except *isSubprocessOf* are locative – *i.e.* all are subsumed by *involves* – so that any pathological process linked to an anatomical structure or process by any chain of these attributes will be considered localised to that structure.

Unusually, GALEN has no notion analogous to "agent" in other systems. Agency is a primary concern of most models of medical record and other information systems in which the GALEN Common Reference Model (CRM) is likely to be used. Therefore it is explicitly left to those systems and excluded from the CRM. There is, however, the notion of "intention" which is required to describe surgical procedures, and of a *VolitionalAct* – a process that has a voluntary intention. However, within the terminology resources, there is no need for a means to identify the actor who will, almost by definition, not be known to it.

Despite its relatively simple structure, this pattern has proved sufficient for extensive modelling both diseases and surgical procedures, including the development of the complete French national surgical procedure classification CCAM [57].

4.3 Diseases

4.3.1 What is a "disease"?

What is a "disease" or "disorder"? What does it mean to say that something is "normal", "abnormal", "pathological" or "physiological"? There are many philosophical definitions[31]. GALEN based its decisions on the pragmatic outcomes required: a sufficient logical approximation that would achieve classifications acceptable to our experts. Required consequences include being able to:

1. Distinguish normal from abnormal anatomy and to list normal anatomical parts, connections, etc. for any structure.
2. Identify entities whose presence was potentially noteworthy in a medical record - *i.e.* "abnormal"
3. Identify entities as in potential need of medical management – *i.e.* as "pathological"
4. Represent the notion of being "abnormal but not pathological" – defined pragmatically as "note-worthy but not in need of medical management"
5. Represent that the presence of some entities is always pathological, *e.g.* a malignant tumour or fracture.

GALEN provides two separate status distinctions intended to address these specific requirements: *normal* vs *nonNormal* and *pathological* vs *physiological* with associated status attributes *hasNormalityStatus* and *hasPathologicalStatus.* In addition it provides stronger versions of *nonNormal* and *pathological*, *intrinsicallyNonNormal* and *intrinsicallyPathological* for those cases in which a category's presence is always *nonNormal* or *pathological*. Using GRAIL's necessary statement mechanism, it is possible to express the following rules:

1. i*ntrinsicallyPathological* → *pathological* → *nonNormal*
2. *intrinsicallyNonNormal* → *nonNormal*

Note that *intrinsicallyNormal* does not imply *normal* nor does *instrinsicallyPhysiological* imply *physiological.* These categories are provided for symmetry and convenience only.

The closest logical approximation to "disease" or "disorder" in GALEN is *PathologicalPhenomenon*, defined as:

Phenomenon *which* *hasPathologicalStatus pathological.*

Combining this notion with the general locative attribute *involves* allows broad disease categories to be defined, *e.g.* "cardiovascular disease" is represented as *CardiovascularPathology* defined as:

Pathologica Phenomenon *which* *involves CardiovascularSystem*

The label *PathologicalPhenomenon* has been explicitly chosen to avoid implying too close a mapping to any natural language phrase such as "disease", "disorder", or

[31] Internal debate within GALEN revealed a surprising diversity of opinion regarding the meaning of both "normal" and "pathological"; the current solution is a pragmatic compromise intended to achieve specific functional goals. Others may prefer alternative labels.

"condition". It has so far proved impossible to reach any consensus on reliable distinctions between such terms.

4.3.2 Causation

Causation, or aetiology, is a critical notion to medical knowledge but surprisingly slippery. GALEN recognises at least two dimensions around causation:

1. Strength of association – from statistical association to physiological cause
2. Timing – temporal relationship between cause and effect (motivated by rheumatic aortitis as a consequence of streptococcal infection but occurring many years later)

Attributes indicating close causal connections are transitive – *e.g. isImmediateConsequenceOf* – whereas attributes indicating loose connections are not – *e.g. isLateConsequenceOf* or *isAssociatedWith*. This is a coarse grained logical approximation for the probabilistic attenuation of causal connection with the length of the causal chain.

Multiple causation gives rise to still more complex issues. Many conditions are defined by their cause, *e.g.* "viral pneumonia", "bacterial meningitis", etc. What is to be done about conditions in which there is more than one cause? Clinicians do not accept the logical inference that "mixed pneumonia" is a kind of "bacterial pneumonia" because they have different implications for management; for the same reason clinicians require the ability to distinguish between a "mixed pneumonia" and a "viral pneumonia complicated by bacterial infection".

GALEN addresses this issue by providing special single-valued child attributes of each causal attribute marked by the naming convention "Specific", *e.g. isSpecificImmediateConsequenceOf.* Using this convention, *ViralPneumonia* is defined as:

Pneumonia <u>which</u> isSpecificImmediateConsequenceOf ViralInfection.

Other dimensions that have been encount ered but not modelled in detail include: a) which of multiple simultaneous effects is considered primary from a clinical point of view; and b) whether an effect is pathophysiologically a direct or indirect consequence of its cause.

5 Application Constructs: Medical Records and Coding Schemes

Two of GALEN's specific objectives are to encapsulate categories so that they can be incorporated into medical records and to provide means of mapping to existing coding and classification schemes. A prerequisite for achieving these objectives is deciding what it is that must be entered into a record, and what should be mapped to a coding scheme. The answers to both questions require additional constructs.

In many electronic medical records, all information must be in the coded expression [8, 10], *e.g.* a code from the Read Clinical Terms [39], SNOMED-CT [75] or earlier schemes such as ICD and its clinical variants [80].

These terminologies have characteristics that are not easy to represent directly in GRAIL or similar formalisms:

1. They include negative as well as positive terms, for example “apyrexia” or “absent pedal pulse”. Many systems that include such terms have no other means of expressing negation.
2. They include complexes of several conditions – *e.g.* A with B without C

To cope with these characteristics, GALEN supports ‘wrapping’ one or more clinical entities in two outer modalities:

Existentiality	presence or absence
ClinicalSituation	A collection of several clinical entities to be recorded together as one “chunk” of clinical information

For example, the expression for “Stomach ulcer with penetration but without haemorrhage” would be:

ClinicalSituation which isCharacterisedBy <
(presence which isExistenceOf StomachUlcer)
(presence which isExistenceOf Stomach Penetration)
(absence which isExistenceOf Haemorrhage)>

For consistency, the wrapping with *ClinicalSituation* and *presence* must be used even when the notion to be represented is just the presence of a single entity, *e.g.*

ClinicalSituation which isCharacterisedBy (presence which isExistenceOf StomachUlcer)

Note that *presence/absence* are not a proper substitute for negation. In the above what is stated logically is the absence of *some Haemorrhage* rather than *any Haemorrhage.* The difference between the semantics of *presence/absence* and true negation must be taken into account when retrieving information from medical records.

However*, presence/absence* works well for mapping to ICD whose “broader than”/ “narrower than” notions work similarly. *ClinicalSituation* therefore provides the basis for mappings to traditonal coding and classification systems such as ICD9/10. The details are beyond the scope of this paper, but key considerations include:

1. The categories in the GALEN Common Reference Model (CRM) do not represent codes directly, rather they are mapped to codes using the indexing methods described in Section 2.3.2. Each ICD, or similar, code is mapped to the most specific corresponding GALEN entity or entities.
2. An ICD, or similar, code may be mapped to more than one GALEN category. Typically this occurs if there is an “includes” or disjunctive clause in the code rubric. In this case it is treated as the disjunction of the GALEN categories to which it is mapped.
3. “Excluding …” clauses in ICD – *e.g.* “hypertension excluding pregnancy” – indicate that a more specific code exists elsewhere in ICD. The indexing method in 1) deals with this automatically. No exceptions to this rule have so far been reported.
4. Any code whose rubric includes “Not otherwise specified” (“NOS”) is mapped to the parent entity with a suitable annotation in the mapping. Likewise for “Not elsewhere classified” (“NEC”) and “Other”
5. All consideration of the rules for handling multiple codes (volume 2 of ICD) are left to external reasoners.

6 Discussion

6.1 Evaluation Against Criteria

In terms of the original criteria of expressiveness, classification and parsimony, GALEN has been sufficiently used in real projects of significant scale to be confident of its expressiveness with respect to either surgical procedures or the clinical information needed to describe the effects and uses of drugs. Surgical procedures were the primary focus of the GALEN-In-USE project, and the tools developed there were subsequently used for the development of the French national surgical classification CCAM [57] and The UK Drug Ontology project [72, 86, 87]. The original use case in clinical information systems has been tested within a limited commercial deployment of a clinical user interface, PEN&PAD/Clinergy [30, 38], based in UK Primary Care.

With respect to classification, cross comparisons have been undertaken with specific subsections of the Clinical Terms Version 3 [62] whilst the entire GALEN ontology has undergone extensive but *ad hoc* manual validation in the course of both GALEN-IN-USE and the Drug Ontology development. These comparisons and quality assurance mechanisms identified errors, but none that led to reconsideration of the basic structure of the ontology.

With respect to parsimony, assessment is more difficult. Constructing an ontology by parsimonious re-use of a deliberately limited set of building blocks inevitably results in increased representational complexity in the way the building blocks are assembled. The question most often raised about GALEN is nearly the converse of parsimony, *i.e.* "Isn't it over engineered?" Would a simpler starting point have been more effective? How much complexity is it worth accepting in return for parsimony? No definitive answer is available. GALEN's response has been to hide the complexity wherever possible. It treats the underlying representation suitable for logical classification as described in this paper as a low level "assembly language" and provides higher level "Intermediate Representations" for authors and users [56, 61].

6.2 Issues with the GRAIL Formalism

Many of the specific details of the Common Reference Model (CRM) follow from limitations of the GRAIL formalism; others are possible because of GRAIL's non-standard features.

The most obvious easily remedied shortcoming is that cardinalities are assigned only to attributes and cannot be specialised when those attributes are used. This results in a proliferation of subattributes that obscure the basic structure. Similarly, disjunction and conjunction of primitives would have helped to clarify the structure and made the intention of notions such as "Phenomenon" clearer. The absence of true negation has not proved a serious problem; its inclusion would bring a major increase in complexity.

That the structural algorithms in the GRAIL classifier are sound but incomplete is well known but has caused little difficulty. The main area of incompleteness can be dealt with relatively easily. Most concern variants on expressions of the form *C1* <u>which</u> *attr1 (C2* <u>which</u> *invAttr1 C1) – e.g.* "a fracture in a limb which is the site of trauma". Such expressions – with cycles of whatever length – have been pragmatically banned from the Common Reference Model. Although legal in modern tableaux algorithm based reasoners, they often cause exponential explosions in classification time.

As described in 2.3.5 and detailed in 3.2.4, GALEN's constructs for inheritance across transitive attributes were originally designed for dealing with part-whole relations, but they have since proved valuable in other contexts. The range of possibilities for achieving the same functionality is much greater today than when GALEN was devised. SEP triples [25, 67] might replace GALEN's constructs in part-whole relations, whilst many of their other functions might also be replaced by constructs in more expressive languages such as OWL. Experimental reasoners supporting "role inclusion axioms" – of which GALEN's *specialisedBy* construct is a subset – have been implemented although they are not yet widely available [29]. An evaluation of the alternatives against defined criteria – both human factors and computational tractability – would be a valuable piece of research. For a preliminary investigation see [42].

Almost uniquely amongst DL based ontologies, GALEN uses both "is part of" and its inverse "has part" (and their subattributes). Both the NCI thesaurus and SNOMED-CT support only "is part of", which is the form required to answer questions such as "What diseases affect the liver or anything that is part of the liver?" Including both "is part of" and "has part" makes classification computationally intractable using now standard tableaux based inference engines, *e.g.* FaCT or Racer. Both "is part of" and "has part" are present in the FMA, but it does not, currently, use DL reasoners. A solution to this limitation in description logic reasoners is urgently required before large biomedical ontologies can be satisfactorily managed using description logics based languages including OWL.[32]

GRAIL is unusual in supporting general inclusion axioms (see Section 2.3.3), but they have proved essential for the ontology. Serendipitously, a side effect of GRAIL's restrictions and GALEN's method of orthogonal taxonomies is that all such axioms are "absorbable" so that they do not have a global impact on the performance of tableaux reasoners [28].

Finally, GRAIL's notation makes it natural to form 'normalised' ontologies with orthogonal taxonomies [44], although the language does not quite force this choice.

6.2.1 Comparison with Other Ontologies

In order to get meaningful comparisons between ontologies, it is first necessary to overcome superficial differences in naming conventions and organisations. For upper ontologies and their modelled extensions this requires careful examination. The most obvious high level comparisons are to DOLCE [17] and BFO [69, 70]. A detailed comparison is beyond the scope of the paper, but some general points follow. GALEN's *Thing* maps very closely to DOLCE's "sortals"; GALEN's disjunction of *GeneralisedStructure* and *GeneralisedSubstance* maps to "Continuant" (BFO) or "Endurant" (DOLCE); *GeneralisedProcess* maps to "Occurrent" (BFO) and "Perdurant" (DOLCE). The major items map smoothly, but there are differences in

[32] The computational issues are independent of philosophical discussions about the comparative status of the two statements, *e.g.* that "normal hands" have five fingers.

the placement of *Collection* and *Feature* that both other ontologies treat as "Continuants". GALEN is intended for use within medical record systems where temporal relations and reasoning are handled external to the ontology; therefore it has only weak notions of time. By contrast, temporal constructs are central to the BFO.

GALEN's *Features* are a reasonable match to DOLCE's "Qualities" and GALEN's *States* to DOLCE's "quale", but neither DOLCE nor BFO have made the distinction between "selectors" and "features" as made in GALEN.

The major difference between the DOLCE and BFO is that DOLCE takes a "cognitivist" view whereas the BFO takes a "realist" view. GALEN's representation is broadly cognitivist. DOLCE makes a distinction between "physical object" and "amount of matter" analogous to GALEN's distinction between *GeneralisedStructure* and *GeneralisedSubstance*. Correspondingly, DOLCE has a role "constitutes" representing the relation between substances and the things made of those substances. GALEN has an equivalent attribute *makesUp/isMadeOf*. "Realists" reject the "constitutes" relation, maintaining that the "physical object" is identical to the "amount of matter" rather than being made of it.

The other obvious comparison is with the anatomy modelling in the Digital Anatomist Foundational Model of Anatomy (FMA) [34, 37, 63]. The FMA, like GALEN, is a domain ontology but confined purely to structural relations. Two groups have independently attempted to reconcile the two ontologies [35, 36, 88-90]. Both met with only limited success, the greatest problem being systematic differences including a) naming conventions; b) the choice of whether or not to reify relations; and c) that GALEN does not enumerate all sanctioned variants, *e.g.* it does not pre-enumerate all possible left and right handed variants of anatomical structures, instead it allows them to be created and classified (post-coordinated) dynamically. A more collaborative attempt at reconciliation dealing with these three issues remains to be performed.

6.3 Outstanding Issues

There are a series of issues that remain outstanding:

- Normative statements, congenital disease, and imputed intentions (See 2.2.7)
- Spatial temporal reasoning and numerical calculations (See 2.3.6)
- Improved handling of the pattern exemplified in "the skin of the hand is a division of the skin of the upper extremity". (See 4.1.1)
- Testing of the consequences of use of SEP triples rather than GALEN's *specialisedBy* axioms (See 6.2)
- How best to take advantage of improvements in description logic and ontology technology now becoming available (See 6.2)

6.4 Summary

GALEN has pioneered the construction of large-scale biomedical ontologies based on description logic. Its experiences illustrate both the advantages and disadvantages of the approach in principle and the limitations of the current state of the art. It provides a set of modelling conventions and patterns that have proved sufficiently robust to be used in practical developments – surgical terminologies, drug information, and data entry systems – which it hopes will continue to provide a useful resource both to developers of biomedical ontologies and as a test corpus for those developing description logic reasoners.

GALEN's pursuit of its combined goals of expressivity, logical classification, and parsimony have led to a complex ontology. However, this complexity can be mitigated for users by intermediate representations and tools. Given adequate support, it has proved accessible and usable. Whether a simpler approach would suffice for future applications, or whether a still more complex approach will be required, remains to be seen.

Acknowledgements

This work supported in part by the European Healthcare Telematics programme in the GALEN, GALEN-IN-USE and SemanticMining projects, by the UK Department of Health Drug Ontology subproject of the Prodigy Project, and by the EPSRC funded HyOntUse project (GR/S44686/01) and the UK JISC funded CO-ODE project, www.co-ode.org. The partners in the GALEN and GALEN-IN-USE projects have contributed extensively. Particular mention is due to Anthony Nowlan who first formulated key parts of what has become the GALEN ontology.

References

1. Artale, A., E. Franconi, and N. Guarino. *Open problems for part-whole relations*. in *International Workshop on Descripition Logics*. 1996. Boston, MA.
2. Artale, A., E. Franconi, and L. Pazzi, *Part-whole relations in object-centered systems: An overview.* Data and Knowledge Engineering, 1996. **20**: p. 347-383.
3. Bateman, J.A., *Upper modelling: a general organization of knowledge for natural language processing.* 1989, USC/Information Sciences Institute.
4. Bittner, T., M. Donnelly, and B. Smith. *Individuals, Universals, collections: On the foundational relations of ontology*. in *International Conference on Formal Ontology and Information Systems (FOIS 2004)*. 2004. Turin, Italy.
5. Borgida, A., *Description logics in data management.* IEEE Transactions on Knowledge and Data Engineering, 1995. **7**(5): p. 671-682.
6. Brachman, R. and H. Levesque. *The tractability of subsumption in frame-based description languages*. in *AAAI-84*. 1984: Morgan Kaufman.
7. Brachman, R.J., et al., *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.
8. Brown, P., M. O'Neil, and C. Price, *Semantic definition of disorders in Version 3 of the Read Codes.* Methods of Information in Medicine, 1998. **37**: p. 415-419.
9. Casati, R. and A.C. Varzi, *Parts and Places*. Parts and Places: The Structurres of Spatial Representation. 1999, Oxford: Clarendon Press.
10. Chute, C., *Clinical classification and terminology: Some history and current observations.* Journal of the American Medical Informatics Association, 2000. **7**(3): p. 293-303.
11. Cimino, J., *Controlled Medical Vocabulary Construction: Methods from the Canon Group.* Journal of the American Medical Informatics Association, 1994. **1**(3): p. 296-197.
12. Cimino, J., *Desiderata for controlled medical vocabularies in the twenty-first century.* Methods of Information in Medicine, 1998. **37**(4-5): p. 394-403.

13. Degen, W., et al. *Formal Ontology in Information Systems (FOIS 2001)*. 2001.
14. Doyle, J. and R. Patil, *Two theses of knowledge representation: Language restrictions, taxonomic classification and the utility of representation services.* Artificial Intelligence, 1991. **48**: p. 261-297.
15. Etherington, D., *Formalising nonmonotonic reasoning systems.* Artificial Intelligence, 1987. **31**: p. 41-85.
16. Evans, D.A., et al., *Position statement: Towards a medical concept representation language.* Journal of the American Medical Informatics Association, 1994. **1**(3): p. 207-217.
17. Gangemi, A., et al. *Sweetening ontologies with DOLCE.* in *European Knowledge Aquisition Workshop (EKAW-2002)*. 2002. Siguenza, Spain: Spring Verlag.
18. Gruber, T.R., *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. 1993, Knowledge Systems Laboratory, Stanford University.
19. Guarino, N. and C. Welty, *An overview of OntoClean*, in *Handbook of Ontologies*, S. Staab and R. Studer, Editors. 2004, Springer Verlag. p. 151-159.
20. Guarino, N. and C. Welty. *Towards a methodology for ontology-based model engineering.* in *ECOOP-2000 Workshop on Model Engineering*. 2000. Cannes, France.
21. Guha, R. and D. Lenat, *Enabling agents to work together.* Communications of the ACM, 1994. **37**: p. 127-142.
22. Haarslev, V. and R. Moeller. *Expresive ABox reasoning with number restrictions, role hierarchies, and transitively closed roes*. in *Proceedings of the Seventh International Conference on Knowledge Representation and Reasoning (KR2000)*. 2000. San Francisco, CA: Morgan Kaufmann.
23. Hahn, U., M. Romacker, and S. Schulz, *How knowledge drives understanding - matching medical ontologies with the needs of medical language processing.* Artificial Intelligence in Medicine, 1999. **15**(1): p. 25-52.
24. Hahn, U., S. Schulz, and M. Romacker, *Part-whole reasoning: a case study in medical ontology engineering.* IEEE Intelligent Systems and their Applications, 1999. **14**(5): p. 59-67.
25. Hahn, U., S. Schulz, and M. Romacker. *Partonomic reasoning as taxonomic reasoning in medicine*. in *Proc. of the 16th National Conf. on Artificial Intelligence & 11th Innovative Applications of Artificial Intelligence (AAAI-99/IAAI-99)*. 1999. Orlando FL: AAAI Press/MIT Press.
26. Hardiker, N.R. and A.L. Rector, *Modeling nursing terminology using the GRAIL representation language.* Journal of the American Medical Informatics Association, 1998. **5**: p. 120-128.
27. Horrocks, I., *Optimising Tableaux Decision Procedures for Description Logics*, in *Computer Science*. 1997, University of Manchester: Manchester. p. 176.
28. 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.
29. Horrocks, I. and U. Sattler., *The decidability of SHIQ with complex role inclusion axioms.* Artificial Intelligence, 2004. **160**(102): p. 79-104.
30. Kirby, J. and A.L. Rector. *The PEN&PAD Data Entry System: From prototype to practical system*. in *AMIA Fall Symposium*. 1996. Washington DC: Hanley and Belfus, Inc.
31. Lenat, D.B. and R.V. Guha, *Building Large Knowledge-Based Systems: Representation and inference in the Cyc Project*. 1989, Reading, MA: Addison-Wesley. 372.

32. Martin, R.F., et al. *Foundational Model of Neuroanatomy:Implications for the Human Brain Project*. in *AMIA Fall Symposium*. 2001. Washington DC.
33. Masolo, C., et al., *WonderWeb Deliveable 18*. 2003, WonderWeb consortium.
34. Mejino, J.L.V. and C. Rosse, *Conceptualization of anatomical spatial entities in the Digital Anatomist Foundational Model.* Journal of the American Medical Informatics Association, 1999(1999 Annual Symposium Special Issue): p. 112-116.
35. Mork, P. and P. Bernstein. *Adapting a generic match algorithm to align ontologies of human anatomy*. in *20th International Conference on Data Engineering*. 2004. Boston: IEEE.
36. Mork, P., R. Pottinger, and P. Bernstein. *Challenges in precisely aligning models of human anatomy*. in *Proceedings of Medinfo 2004*. 2004. San Francisco, CA: IMIA.
37. Neal, P.J., L.G. Shapiro, and C. Rosse, *The Digital Anatomist structural abstraction: a scheme for the spatial description lf anatomical entities.* Journal of the American Medical Informatics Association, 1998((Fall Symposium Special Issue)): p. 423-427.
38. Nowlan, W.A., *Clinical workstation: Identifying clinical requirements and understanding clinical information.* International Journal of Bio-Medical Computing, 1994. **34**: p. 85-94.
39. O'Neil, M., C. Payne, and J. Read, *Read Codes Version 3: A user led terminology.* Methods of Information in Medicine, 1995. **34**: p. 187-192.
40. Odell, J.J., *Six different kinds of composition.* Journal of Object Oriented Programming, 1994. **5**(8): p. 10-15.
41. Padgham, L. and P. Lambrix. *A framework for part-of hierarchies in terminological logics*. in *KR-94*. 1994.
42. Rector, A. *Analysis of propagation along transitive roles: Formalisation of the GALEN experience with medical ontologies*. in *2002 International Workshop on Description Logics (DL2002)*. 2002. Toulouse France: CEUR-Proceedings 53.
43. 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.
44. Rector, A. *Modularisation of domain ontologies Implemented in description logics and related formalisms including OWL*. in *Knowledge Capture 2003*. 2003. Sanibel Island, FL: ACM.
45. Rector, A., *Thesauri and formal classifications: Terminologies for people and machines.* Methods of Information in Medicine, 1998. **37**(4-5): p. 501-509.
46. Rector, A., et al., *The GRAIL concept modelling language for medical terminology.* Artificial Intelligence in Medicine, 1997. **9**: p. 139-171.
47. Rector, A., et al. *Scale and Context: Issues in ontologies to link health- and bio-Informatics*. in *AMIA Fall Symposium*. 2002. Austin Texas: Hanley and Belfus, Philadelphia.
48. Rector, A., et al. *Making sound re-usable terminology practical: The GALEN approach*. in *Towards and Electronic Health Care Record, Europe*. 1998. London: Medical Records Institute, Newton Mass.
49. Rector, A., et al., *A Terminology Server for Medical Language and Medical Information Systems.* Methods of Information in Medicine, 1995. **34**: p. 147-157.
50. Rector, A., A. Taweel, and J. Rogers. *Models and inference methods for clinical systems: A principled approach*. in *Medinfo*. 2004. San Francisco: North Holland.
51. Rector, A., et al., *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.

52. Rector, A.L., *Clinical Terminology: Why is it so hard?* Methods of Information in Medicine, 1999. **38**: p. 239-252.
53. Rector, A.L. *Defaults, context and knowledge: Alternatives for OWL-Indexed Knowledge bases*. in *Pacific Symposium on Biocomputing (PSB-2004)*. 2004. Kona, Hawaii: World Scientific.
54. Rector, A.L. *The Interface between Information, Terminology, and Inference Models*. in *Tenth World Conference on Medical and Health Informatics: Medinfo-2001*. 2001. London, England.
55. Rector, A.L., et al. *Interface of inference models with concept and medical record models*. in *Artificial Intelligence in Medicine Europe (AIME)*. 2001. Cascais, Portugal: Springer Verlag.
56. Rector, A.L., 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): p. 229-242.
57. Rodrigues, J.M., et al. *Galen-In-Use: An EU Project applied to the development of a new national coding system for surgical procedures: NCAM*. in *Medical Informatics Europe '97*. 1997. Porto Carras, Greece: IOS Press.
58. Rogers, J. and A. Rector. *The GALEN ontology*. in *Medical Informatics Europe (MIE 96)*. 1996. Copenhagen: IOS Press.
59. Rogers, J. and A. Rector, *GALEN's model of parts and wholes: Experience and comparisons.* Journal of the American Medical Informatics Association, 2000((Fall symposium special issue)): p. 819-823.
60. Rogers, J., et al. *From rubrics to dissections to GRAIL to classifications*. in *Medical Informatics Europe (MIE-97)*. 1997. Thesalonika, Greece: IOS Press.
61. Rogers, J.E., *Development of a methodology and an ontological schema for medical terminology*, in *School of Medicine, Dentistry, Nursing and Pharmacy*. 2004, University of Manchester. p. 185.
62. Rogers, J.E., et al., *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): p. 845-849.
63. Rosse, C., I.G. Shapiro, and J.F. Brinkley, *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): p. 820-824.
64. Schank, R.C. and R. Abelson, P, *Scripts, Plans, Goals, and Understanding*. 1977, Hilsdale, NJ: Lawrence Erlbaum Associates.
65. Schulz, S. and U. Hahn. *Mereotopological reasoning about parts and (w)holes in bio-ontologies*. in *Formal Ontology in Information Systems (FOIS-2001)*. 2001. Ogunquit, ME: ACM.
66. Schulz, S. and U. Hahn. *Parts, locations, and holes - Formal reasoning about anatomical structures*. in *Artificial Intelligence in Medicine Europe (AIME-2001)*. 2001. Cascais, Portubal: Springer.
67. Schulz, S., U. Hahn, and M. Romacker. *Modeling anatomical spatial relations with description logics*. in *AMIA Fall Symposium (AMIA-2000)*. 2000. Los Angeles, CA: Hanly & Belfus.
68. Shahar, Y., et al. *A problem-solving architecture for managing temporal data and their abstractions*. in *Workshop on Implementing Temporal Reasoning, AAAI-92*. 1992. San Jose, CA.
69. Smith, B. *The basic tools of formal ontology*. in *Formal Ontology in Information Systems (FOIS)*. 1998. Amsterdam: IOS Press (Frontiers in Artificial Intelligence and Applications).

70. Smith, B. *The logic of biological classification and the foundations of biomedical ontology*. in *10th International Conference in Logic Methodology and Philosophy of Science*. 2004. Oviedo Spain: Elsevier-North-Holland.
71. Solomon, W. and H. Heathfield. *Conceptual modelling used to represent drug interactions*. in *Twelfth International Congress of the European Federation for Medical Informatics, MIE-94*. 1994. Lisbon, Portugal.
72. Solomon, W., et al., *A reference terminology for drugs.* Journal of the American Medical Informatics Association, 1999((Fall Symposium Special Issue)): p. 152-155.
73. Sowa, J., *Knowledge Representation*. 1999: Morgan Kaufmann.
74. Staab, S. and A. Maedche. *Ontology engineering beyond the modeling of concepts and relations*. in *ECAI 2000. 14th European Conference on Artificial Intelligence; Workshop on Applications of Ontologies and Problem-Solving Methods*. 2000.
75. Stearns, M., et al. *SNOMED clinical terms: overview of the development process and project status*. in *AMIA Fall Symposium (AMIA-2001)*. 2001: Henley & Belfus.
76. The Gene Ontology Consortium, *Creating the gene ontology resource: design and implementation.* Genome Research, 2001. **11**: p. 1425-1433.
77. The Gene Ontology Consortium, *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**: p. 25-29.
78. Touretzky, D., *The Mathematics of Inheritance Systems*. 1986, Los Altos, CA: Morgan Kaufmann.
79. Tuttle, M.S., *The position of the canon group: A reality check.* Journal of the American Medical Informatics Association, 1994. **1**(3): p. 298-299.
80. United States Center for Health Statistics, *International Classification of Diseases, Ninth Revision with Clinical Modifications*. 1980, Washington, DC: The Center.
81. Uschold, M. and M. Gruninger, *Ontologies: principles, methods and applications.* Knowledge Engineering Review, 1996. **11**(2).
82. Varzi, A.C., *Mereological commitments.* Dialectica, 2000. **54**: p. 283-305.
83. Welty, C. and N. Guarino, *Supporting ontological analysis of taxonomic relationships.* Data and Knowledge Engineering, 2001. **39**(1): p. 51-74.
84. Winston, M., R. Chaffin, and D. Hermann, *A taxonomy of part-whole relations.* Cognitive Science, 1987. **11**: p. 417-444.
85. Woods, W., *What's important about knowledge representation?* IEEE Computer, 1983. **16**: p. 22-29.
86. Wroe, C. and J. Cimino. *Using openGALEN techniques to develop the HL7 drug formulation vocabulary*. in *American Medical Informatics Association Fall Symposium(AMIA-2001)*. 2001.
87. Wroe, C., et al., *Inheritance of drug information.* Journal of the American Medical Informatics Association, 2000(Annual Symposium Special Issue): p. 1158.
88. Zhang, S. and O. Bodenreider. *Comparing associative concepts among equivalent concepts across ontologies*. in *Medinfo-2004*. 2004. San Francisco, CA: IMIA, IOS Press.
89. Zhang, S. and O. Bodenreider. *Investigating implicit knowledge in ontologies with application to the anatomical domain*. in *Pacific Symposium on Bioinformatics*. 2004: Bioinformatics.
90. Zhang, S., P. Mork, and O. bodenreider. *Lessons learned from aligning two representations of Anatomy*. in *First international workshop on formal biomedical knowledge representation*. 2004. Whistler, Canada.
91. Zweigenbaum, P., et al., *Issues in the structuring and acquisition of an ontology for medical language understanding*. Methods of Information in Medicine, 1995. **34**(1/2): p. 15-24.