“Patterns, Properties and Minimizing Commitment: Reconstruction of the GALEN Upper Ontology in OWL

Alan L Rector and Jeremy Rogers

Department of Computer Science, University of Manchester, Manchester M13 9PL, UK
email: rector@cs.man.ac.uk; Tel +44-161-275-6188; FAX +44-161-275-6204

Abstract. The GALEN upper ontology has been used over a period of more than a decade for representing biomedical concepts including anatomy, drugs, diseases, signs and symptoms. A reconstruction of the key features of the ontology in OWL-DL is presented along with the underlying principles – that “all distinctions should make a difference”, that “nonessential distinctions should be deferred”, and that “the ontology should be ‘normalized’”.

Introduction

The upper ontology and patterns used in the GALEN models for biomedical applications have stood the test of time within the project and its extensions into drug information and bioscience [9, 11][1]. The GALEN domain level ontologies embody the principles of “normalization” and the pattern for value types[2] described in [5]. However, the GALEN upper ontology is based on a quite different set of principles. In general, the greater expressivity of OWL and modern description logics compared with the original implementation language, GRAIL [6], makes it possible to reformulate these principles more clearly. However, the lack of inheritance along transitive properties in OWL (the `propagates_via`[3] construct[4] from GRAIL) means that an adaptation of Schulz and Hahn’s mechanism of ‘SEP triples”[2-4] must be used to reformulate the part-whole relations. Efforts to harmonise with the Digital Anatomist Foundational Model of Anatomy[10][4] have brought further modifications. The terminology for some high level abstractions has been influenced by Ontoclean[1].

Basic Principles

The design of the GALEN ontology has been based from the beginning on a two high level principles:

---

[1] http://www.opengalen.org
[2] Termed by the Semantic Web Best Practice Group: “value partitions”
[3] Literally in GRAIL “`refined_along`” and “`specialized_by`”
[4] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

- That the ontology should be implemented in a logical formalism permitting automatic classification; today, in OWL-DL or a plausible extension.
- That all distinctions should make a difference in engineering terms to the representation of information. When faced with alternatives, the questions should be: “What will the consequences be?” “What can or cannot be said in each case?”

As with most ontologies, the GALEN ontology is divided into an upper ontology consisting of entities that are largely domain independent and a series of domain level modules concerning of specific biomedical areas.

For the upper ontology the key principles are:

- That the upper ontology should consist of a set of mutually disjoint notions – *e.g.* what would now be termed *Continant* vs *Occurrent*, *Mass* vs *Discrete*, etc.[5]
- That the distinctions should be capable of being made reliably by authors without undue argument.
- That the patterns for analogous constructs should be consistent.
- That the patterns should support the part-whole relations necessary for describing anatomy and its interrelations with procedures and disorders.
- That, where possible, decisions should be deferred so as to minimize conflicts when being reconciled with other ontologies.

In addition, two feature of GRAIL’s expressivity greatly affect the outward form of the GALEN ontology:

- That GRAIL attaches all cardinality information to unqualified properties[6], which leads to a proliferation of subproperties whose only role is to govern cardinality.
- That GRAIL lacks disjunction, which results in asserted primitives such as ‘Phenomenon’ that obscure the basic structure of the upper ontology.

The remainder of this paper describes the principles behind the reconstruction of GALEN’s upper level ontology in OWL-DL[7]. The goal of this reconstruction is to reach the level of starting domain concepts – for biology – organisms, organs, cells, tissues, etc. As a first test we have represented the proposed “Standard Entry Anatomy List”[8] being developed by SOFG in the revised structure. Table 1 at the end of the paper summarises the main distinctions and their rationales.

---

[5] In GALEN, *GeneralisedSubstanceOrStructure* vs *GeneralisedProcess, Structure* vs *Substance, Physical* vs *Abstract,*

[6] Termed “attributes” in GRAIL.

[7] An example implementation in OWL-DL is provided at www.co-ode.org/ontologies/sample-top-bio/. and/or www.cs.man.ac.uk/~rector/ontologies/sample-top-bio/

[8] http://www.sofg.org/sael/

Reconstruction of the GALEN Upper Ontology

Self-Standing[9] vs Refining Entities

Entities that are fully dependent on other entities – *e.g.* size, shape, malignancy, severity, etc. – are termed "*Refining*". All others are "*Self Standing*" [10]. Refining entities are further subdivided into "*Modifiers*" that represent "qualities" that might themselves be further refined or described – *e.g.* severity, size, body temperature – and "*Selectors*" which merely indicate one of a set – *e.g.* left and right laterality; upper, middle and lower position; number of fingers, vertebrae, etc.

Modifiers are always "reified" – *i.e.* there is a class for each "quality". Each individual can have exactly one of each kind of quality – *e.g.* one height, one body temperature, one severity, etc. However, each quality can have one value from each of an indefinite number of value partitions – *e.g.* level, trend, quantitative value, etc. This allows the representation of qualities such as "Temperature that is elevated, rising, and has a quantitative value of 39°C."[11]

Distinctions and patterns for Self-Standing Entities

There are three main distinctions in GALEN, if rephrased in modern terminology:

- *Continuants vs Occurrents*[12] or roughly "things vs processes". The key pattern is that "Things participate in processes" in various ways. GALEN is concerned with how diseases and procedures "act on" things or have things as their "actors"[13].
- *Discrete vs Mass Entities*. The key pattern is "Things are made of stuff". Example: "Organ is made of tissue"; "Nail is made of steel" Most existing ontologies maintain this distinction, although the property is usually termed "is_constituted_of/constitutes". However, for occurrents, the distinction is often difficult in practice *e.g.* between "digestion" and "digestion of a meal". After much debate, GALEN chose not make the mass/discrete distinction for occurrents. In the reconstruction, this decision is deferred a) by not making the distinction exhaustive, and b) by not specifying whether most occurrents are mass or discrete. The issue of how to treat mass entities that have structure, such as biological tissues, is highly controversial within the SAEL community and in reconciliation with the Digital Anatomist Foundational Model of Anatomy, but it has posed little difficulty within GALEN's applications.
- *Physical vs non_Physical*[14] and within physical *Material vs non_Material*. GALEN, being concerned with biomedicine, is concerned primarily with physical

---

[9] In GALEN this distinction is made as between 'Modifiers' and everything else.

[10] The neutral term "self_standing" has been chosen to allow discussion of whether or not this notion is really equivalent to the notion of "independent" entity as used by various other authors.

[11] Representation of this pattern in DAML+OIL is straight forward. The absence of "qualified cardinality constraints" in OWL makes its representation unsatisfactory. GRAIL also lacked qualified cardinality restrictions leading to a proliferation of properties of the form hasTemperatureFeature, hasHeightFeature, etc.

[12] "GeneralisedStructure or GeneralisedSubstance" vs "GeneralisedProcess" in GALEN

[13] GALEN used "isFunctionOf". It originally used 'agent of' but this conflicted with linguistic usage. .

entities. Essentially all critical entities in GALEN are physical. GALEN indicates a "topology" for all physical structures. "Hollow structures" define "spaces"; "spaces" can contain other structures and can be subdivided into "subspaces". The reconstruction adopts the FMA terminology of "material" and "non-material", where "non-material" includes abstract "lines" and "joins" as well as "spaces". "Non-material entities" are "defined by" "material entities" or are subdivisions of other "non-material entities".

Three further issues are dealt with in the upper ontology:

- *Biological vs non-Biological.* A convenience in a biomedical ontology.
- *Phenomenon.* A prominent and peculiar structure in the GALEN upper ontology is the entity "*Phenomenon*". The notion of "*Phenomenon*[15]" was created, in effect, as a disjunction of all things that could have a pathological status – *i.e.* be normal, non-normal or pathological, *i.e.* "diseased" [9]. *"Phenomena"* include structures (*e.g.* tumours), substances (*e.g.* pus), altered processes (*e.g.* heart failure), qualities (*e.g.* elevated temperature) and processes (*e.g.* inflammation) but not selectors, values, or magnitudes. In OWL the domain of the corresponding property – `has_abnormality_status` – can be an anonymous disjunction, eliminating the need for a named entity. This greatly clarifies the structure of the ontology.
- *Disease & "Clinical Situation":* In many contexts a "disease" is a complex of a primary phenomenon and other associated or resultant phenomena. In addition, since GRAIL lacks negation, the notion of "presence" or "absence" has to be modeled explicitly. As a result, for most purposes, the entity in GALEN most closely approximating "disease" is a complex of the presence one or more pathological phenomena and the presence or absence of other phenomena. GALEN's class for such complexes is "*Clinical Situation"*. The greater expressivity of OWL makes the scheme simpler, but the notion of Clinical Situation, or "syndrome" is retained.
- *Dualities:* Many continuants (physical structures) are always the outcome of corresponding occurrents (processes) – *e.g.* "ulcer" and "ulceration" or "tumor" and "proliferation". Often the process and structure share the same name, *e.g.* "erosion". A consistent representation for structure-process duals must be chosen: either one is always defined in terms of the other or they are both taken as "aspects" of a complex. GALEN almost always choses to define structures as the outcome of processes. Experiments with both this solution and the alternative of making both "aspects" of a complex have been conducted in the reconstruction. The consequences of the two alternatives are still being explored.

---

[14] GALEN "Abstract" – dropped in the reconstruction to avoid confusion with the many other usages of the label "abstract"

[15] Originally "ClinicalPhenomenon" but later shortened despite the obvious clash with usages in other contexts.

Reconstruction of the GALEN partonomy relationships

GALEN supports a series of inferences about partonomy that require distinguishing several cases. The principles are set out below and summarised in Table 2.

- *Principles applying to transitive partitive properties and containment*
  - *A disease of the part is a disease of the whole* (unless specifically stated to the contrary). In GALEN, this is achieved by an axiom stating that the property linking diseases to anatomical sites, `has_locus`[16], `propagates_via` the property `is_part_of`. To achieve this in OWL, which lacks the `propagates_via` construct, requires rewriting disease and procedure definitions using a variant of the 'SEP triple' mechanism[2] *e.g.* "Heart disease" is defined as "a disease of the heart or any of its parts".
  - *Diseases, actions on, and functions of the whole must be distinguished from diseases of the whole or its parts.* For example, there must be a way to ensure that "Removal of a lobe of liver" is not a kind of "Removal of liver". In GALEN this is done awkwardly through special sub-properties. In the reconstruction using adapted SEP triples, there are distinct entities for the whole and the parts. The distinction can therefore be made directly.
  - *Functional partonomy does not imply structural partonomy nor vice versa*. For example, the glands of the endocrine system are not structural parts of any one thing.
  - *Mass entities are portions of other mass entities – e.g.* "water is a portion of blood".[17] In addition, *the portion a part is a portion or the corresponding portion of the whole – e.g.* "the blood in the heart is a portion of the blood in the circulatory system". The second part of this rule cannot be implemented within the DL fragment, but the distinctions necessary to support it are retained pending further advances in representation technology, *e.g.* SWRL[18].
  - *Containment does not imply partonomy – e.g.* "the brain is contained in the (cavity of the) skull", and "the heart is contained in the thorax"; but neither organ is part of the containing structure. Things contained in parts are contained in the whole, but parts of contained things are not parts of the container.
  - *Discrete entities are constituted of mass entities – e.g.* organs are constituted of tissue.
  - *Distinct parts of subdivisions are distinct parts of the whole but not vice versa.* Examples of distinct parts[19] include valves, ligaments, muscles, nerves, etc. as opposed to generic regions such as "left side" or "lobe". For example, "lobe of the liver" is not a "subdivision of the body" whereas the "cusp of a heart valve" is a "component of the heart". This distinction is awkward to achieve in OWL without the `propagates_via` axiom, and it is ignored in the reconstruction.

[16] GALEN literally uses "has_location" but this has caused confusion in more general audiences with the purely locative property of location. SNOMED and several other schemes use has_site.

[17] GALEN "makes up in part"

[18] http://www.w3.org/Submission/SWRL/

[19] "Components" in GALEN

  - *Layers of subdivisions are layers of the whole* – e.g. “the skin of the hand is a kind of the skin of the upper extremity.” This is an approximation. It would be more correct to say that the “skin of the hand is a subdivision of the skin of the upper extremity”, but the necessary rule cannot be expressed within the DL fragment of first order logic.
- *Non-transitive – partonomy across boundaries of scale*
  - *Discrete entities at one granularity form collectives or “multiples” at the next that act as mass entities, e.g.* collectives of molecules form water. Boundaries between levels of granularity are not fixed but rather determined case by case by whether the information concerns the collective as a whole or the discrete entities taken individually.
- *Non-partitive locative properties*
  - *Connection, branching, adjacency, etc. do not imply part-whole relations*
- Locus and Location[20]
  - *Diseases have a “locus” which may not be strictly spatial.* For example, a “Disease of the endocrine system” where the “endocrine system” has no simple spatial location. Clinical usage is so heavily influenced by the notion of “locus” or “site” that it is almost always given special status. In particular “locus” is propagated by functional as well as structural partonomy.
  - *Physical location is implied by physical partonomy and containment*, but not by functional partonomy or locus.

Summary and open questions

The major distinctions in the reconstructed GALEN ontology, examples, and their engineering consequences are given in Table 1. Much of the success of the ontology – the fact that users could be taught to use it in a few days – stemmed in GALEN from the use of an Intermediate Representation which relaxed many of constraints [8]. That the constraints could be relaxed for users stemmed from the fact that they could be inferred from distinctions in the upper ontology. The current reconstruction is clearer. However, as in GALEN, it is intended that the upper ontology be largely hidden from users.

The three most problematic areas remaining are: a) process-outcome duals, b) the status of mass entities that have structure – tissues, or more generally, “materials” – and c) the partonomic principles that cannot be expressed within the DL fragment of first order logic. All require awkward “work arounds”. The “work arounds” in turn require support in tools – even for experienced knowledge engineers, let alone domain experts.

This paper has concentrated on structure – or more broadly, on “continuants”. The GALEN ontology also includes processes (“occurrents”) but has explicitly confined itself to biology and related physical entities. It contains only limited notions of agency, responsibility, or other non-physical abstractions. Likewise, GALEN

---

[20] The GALEN “has_location” and “Locative property” have rough equivalent in the reconsultrction are “has_locus” and “has_lcation”.

excludes entities that are expected to be dealt with by mechanisms other than simple subsumption reasoning. Most importantly numbers, other mathematical constructs and most temporal constructs are excluded. GALEN's goal has been limited to providing a common structure for clinical and biomedical entities and the relations amongst them. It is assumed that it will be used as part of larger systems that support numeric and temporal data by other means, for example, in conjunction with models of healthcare records and clinical decision support [7, 8].

Acknowledgements

This work supported in part by the CO-ODE grant from the Joint Information Services Committee (JISC) and the HyOntUse GR/S44585 from the UK EPSRC and by the Semantic Mining Network of Excellence funded by the EC.

References

1. Guarino, N. and Welty, C., Towards a methodology for ontology-based model engineering. in *ECOOP-2000 Workshop on Model Engineering*, (Cannes, France, 2000).
2. Hahn, U., Schulz, S. and Romacker, M., 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)*, (Orlando FL, 1999), AAAI Press/MIT Press, 271-276.
3. Hahn, U., Schulz, S. and Romacker, M. Part-whole reasoning: a case study in medical ontology engineering. *IEEE Intelligent Systems and their Applications*, *14* (5). 59-67.
4. Rector, A., Analysis of propagation along transitive roles: Formalisation of the GALEN experience with MedicalOntologies. in *2002 International Workshop on Description Logics (DL2002)*, (Toulouse France, 2002), CEUR-Proceedings 53, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-53/Rector-DL2002-propagates-via-final.ps.
5. Rector, A., Modularisation of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. in *Knowledge Capture 2003*, (Sanibel Island, FL, 2003), ACM, 121-128.
6. Rector, A., Bechhofer, S., Goble, C., Horrocks, I., Nowlan, W. and Solomon, W. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, *9*. 139-171.
7. Rector, A.L., Johnson, P.D., Tu, S., Wroe, C. and Rogers, J., Interface of inference models with concept and medical record models. in *Artificial Intelligence in Medicine Europe (AIME)*, (Cascais, Portugal, 2001), Springer Verlag, 314-323.
8. Rector, A.L., Zanstra, P.E., Solomon, W.D., Rogers, J.E., Baud, R., Ceusters, W., W Claassen, Kirby, J., Rodrigues, J.-M., Mori, A.R., Haring, E.v.d. and Wagner, J. Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine. *IEEE Transactions on Information Technology in BioMedicine*, *2* (4). 229-242.
9. Rogers, J. and Rector, A., The GALEN ontology. in *Medical Informatics Europe (MIE 96)*, (Copenhagen, 1996), IOS Press, 174-178.
10. Rosse, C., Shapiro, I.G. and Brinkley, J.F. The Digital Anatomist foundational model: Principles for defining and structuring its concept domain. *Journal of the American Medical Informatics Association* (1998 Fall Symposium Special issue). 820-824.
11. Wroe, C. and Cimino, J., Using openGALEN techniques to develop the HL7 drug formulation vocabulary. in *American Medical Informatics Association Fall Symposium(AMIA-2001)*, (2001), 766-770.

Distinction	Example	Test Questions	Significance
Top Self_standing vs Refining	Hand vs left	Can you talk about it on its own? Is the modified thing a kind of the original?	*Self standing entities* form the skeleton of the domain ontology. Only self_standing entities can be the object or value of *relational properties*
Within Refining Selecting vs Modifying	Left (hand) vs Temperature (of hand)	Can the refiner be changed without changing the refined?	Modifiers should always be reified; reifying selectors serves no purpose
Within Self_standing Continuants vs Occurrents	Organism vs Metabolism	Does it continue (endure) through time or does it occur over time?	Entities can only *participate in Occurrents*.
Mass vs Discrete	Tissue vs Organ	Do you measure it or count it?	*Discrete entities* are *constituted* of *Mass entities*
Physical vs nonPhysical.	Organ vs Idea	Does it occupy/occur in space?	GALEN is primarily concerned with physical entities. A placeholder
Biological vs nonBiological	Heart vs Prosthesis	Is it an organism or from an organism?	A primary organizational principle for users
Within Physical Material vs nonmaterial	Chest vs Chest Cavity	Is it a material thing or something marked out by a material thing?	*Material entities* (and their junctions and featuress) *define non material entities*[21]
Within Continuants Complexes vs Noncomplexes[22]	Systems vs Organs, Multiples of Cells vs a Cell.	Does the complex have properties collectively distinct from its individual members?	Only *Complexes* can have *members*.
Within Complexes Groups vs Multiples	System of Organs vs Cells in a tissue	Is the complex a discrete or mass entity?	Boundaries of Granularity/Scale. *Multiples* of *descrete entities* at one scale are *mass entities* at the next.

Table 1: Key top level distinctions

Property	Usage	Transitivity
has_location	Generic top property	T
has_locus	Links disorders to ‘sites’ which can be distributed	T
has_physical_location	Primary physical location – any two physical entities	T
is_part_of	Generic partonomic relation	T
is_structural_part_of	Structural physical part-hood, heavily influenced by anatomy	T
is_defined_by	Nonmaterial by material entities, *e.g.* Cavities by Cavitated organ	T
is_portion_of	Mass entity of mass entity: *e.g.* water of blood	T
constitutes	Mass entity of discrete entity; *e.g.* tissue of organ	T
is_subdivision_of	Generic part of defined part, *e.g.* lobe of liver	T
is_distinct_part_of	Defined part of defined part, *e.g.* valve of heart	T
is_functional_part_of	Part that contributes to functioning, *e.g.* heart to circulatory system	T
is_member_of	Generic parent of relations of things to complexes	¬T
is_grain_of	Discrete to mass, *e.g.* cell to tissue; marks granularity boundary	¬T
is_in_group_of	Discrete collections, *e.g.* heart to circulatory system	¬T
is_contained_in	Material in nonMaterial, *e.g.* Brain in cavity of skull[23].	T
is_connected_to	Discrete physical to discrete physical: Aorta is connected to Heart.	¬T[24]

Table 2: Major locative and partitive properties

---

[21] Dealt with in GALEN using selector hasTopology.

[22] NB: The issue of whether or not complexes are physical deliberately deferred.

[23] GALEN simplified by allowing one material entity to be contained directly in another

[24] GALEN uses connection in the narrow sense of ‘having a connection to’. Most connections are in fact reified.