Scale and Context: Issues in Ontologies to link Health- and Bio-Informatics

Alan Rector MD, Jeremy Rogers MB BS, Angus Roberts, BSc MSc, Chris Wroe MB BS
Medical Informatics Group, Department of Computer Science, University of Manchester, UK
http://www.cs.man.ac.uk/mig/ rector@cs.man.ac.uk

Abstract

Bridging levels of scale and context are key problems for integrating Bio- and Health Informatics. Formal, logic-based ontologies using expressive formalisms are naturally "fractal" and provide new methods to support these aims. The basic notion of composition can be used to bridge scales; axioms can be used to carry implicit information; specific context markers can be included in definitions; and a hierarchy of semantic links can be used to represent subtle differences in point of view. Experience with *Open*GALEN, the UK Drug Ontology and new experiments with the Gene Ontology and Foundational Model of Anatomy suggest that these are powerful tools provide practical solutions.

INTRODUCTION

Two key problems in linking health- and bio-informatics are:

- *Bridging levels of scale* – between the atomic and molecular scales typical of bioinformatics, and the organ, organism, and even organisational scales typical of clinical medicine and health informatics.
- *Expressing differences in context* – between the views of different professions, between homologous information in different organisms or stages of development and between normal and abnormal anatomy and physiology.

Logic based, rigorous ontologies – formal explicit specifications of shared conceptualisations [1, 2] – have usually been thought of in health informatics primarily as ways of organising terminologies, *e.g.* SNOMED-RT/CT[1] [3] or the *Open*GALEN[2] resources [4]. However, they also provide powerful means of linking differing ontologies – for example the different scales to integrate health- and bio-informatics – and for expressing contextual information and views – *e.g.* the strict structural view of the anatomist or the more pragmatic view of the clinician.

The fundamental principles for managing scale and context are based on:

- Composite concepts, made up of subconcepts from two or more ontologies, can bridge between those ontologies.
- Markers for context can be included directly within such composite concepts.
- Variant semantic links can be used in different contexts.
- The reasoner or 'classifiers' that manipulate these ontologies can identify equivalences, subconcept relations (subsumption), and inconsistencies.

A major advantage of formal ontologies for such tasks is that they are naturally "fractal" – they allow patterns to be repeated at progressively finer levels of detail and smaller scales. Unlike many database techniques, decisions concerning specific levels of detail or structure need not be fixed. Structures can be elaborated and evolve as required.

These techniques are of increasing importance because rigorous formal ontologies are becoming widespread in the biomedical community and underpin much of the new work on the Semantic Web by the W3C[3] including its new language – known variously as "OIL", "DAML+OIL" and "OWL"[4] [5]. (Here we refer to it as "DAML+OIL/OWL".) The bioinformatics community, in particular, is playing a leading role. Examples in this paper will be taken both from work with an older ontology language, GRAIL, and the emerging standard DAML+OIL/OWL, and from experiments in clinical medicine in *Open*GALEN [6] using the Digital Anatomist Foundational Model of Anatomy [7], and the Gene Ontology[5] .

THE BASIC PRINCIPLE: LINKING INDEPENDENT ONTOLOGIES

Fundamentals features of logic based ontologies

The terminology used differs between knowledge representation languages, but most logic-based ontologies consist of at least:

- *Primitive concepts* – usually placed in a skeleton hierarchy and additionally *described* by *necessary* conditions expressed as boolean combinations of other primitives, descriptors or composite concepts.
- *Composite concepts* – *defined* by *necessary and sufficient* conditions expressed in the same way.
- *Properties*[6] – which express the semantic links between concepts and can themselves be placed in a hierarchy,

---

[1] www.snomed.org
[2] www.opengalen.org
[3] www.semanticweb.org
[4] www.daml.org/2001/03/daml+oil-index
[5] www.geneontology.org
[6] Sometimes termed "slots", "attributes", "roles" or "semantic links'

declared equal to the inverse of other properties, and can be functional or transitive.

- *Descriptors*[7] – property-concept pairs qualified by either "some" (∃) or "only" (∀)[8], *e.g.* *"hasLocation-some-Leg", "hasLaterality only left"* etc.
- *Axioms* – which declare concepts either to be disjoint or to imply descriptors or other concepts. (The expressive power of axioms differs greatly amongst ontology languages.)
- *A reasoner* – which infers further superconcept-subconcept (subsumption) relations and equivalences on the basis of the definitions, descriptions and axioms and can check for their logical consistency.

If a tractable reasoning algorithm is known for the set of features in an ontology language we describe that language as 'computational'. Typical computational ontology languages include the description logic used in SNOMED-RT/CT, GRAIL used in OpenGALEN, and DAML+OIL/OWL.

We term that part of the hierarchy consisting only of the primitive concepts its "primitive skeleton". An ontology is "normalised" if its primitive skeleton a) consists of pure single hierarchies of disjoint concepts – *i.e.* each primitive concept has exactly one primitive parent; b) conforms to basic notions of soundness outlined by Guarino [8].

For brevity and clarity, we use an informal notation. Quantification in ontologies is often an important source of confusion so, where appropriate, we include the qualifier "some" routinely to indicate existential quantification[9] even though it makes the English awkward. Correspondingly we use "the" or omit the qualifier entirely to indicate universal quantification of a single-valued (functional) property or slot. In this way, all notations can be rigorously and directly translated into the formal notation of DAML+OIL/OWL.

Linking normalised ontologies

Composite concepts can be thought of as linking normalised subontologies. For example, in SNOMED-RT/CT and GALEN, subontologies of anatomy, disease process, micro-organisms, etc. are linked by standard relations to form sets of compositional statements – *e.g.:*

"Inflammation of some lung caused by some infection with some pneumoccocus"

In this familiar example, notions from separate subontologies – morphology, anatomy, and microorganisms – are linked logically into a concept bridging those ontologies. This same approach works equally well when the ontologies are from different scales – *e.g.:*

"SNPolymorphism of CFTRGene causing some Defect in some MembraneTransport of ChlorideIon causing some Increase in the Viscosity of some Mucus in some CysticFibrosis"

Note that in this example, as in most cases, most of the concepts come from well separated ontologies, each appropriate to a specific scale: *SNPolymorphism* and *Gene* from the cellular, *ChlorideIon* from the molecular, *Viscosity* and *Mucus* from the macro, and *CysticFibrosis* from the disease section of the macro scale. The one potentially troublesome notion is "transport" which has been disambiguated as *MembraneTransport*, a notion which clearly comes from the ontology of the cellular scale.

This clear separation is the key requirement for composition across ontologies to work as expected – *i.e.* that the ontologies to be linked are independent and normalised. The individual ontologies may be as internally complex as required, but for the process of linkage by composition to work, they must not overlap. Otherwise, meanings may be unintentionally confused.

When the ontologies to be linked already overlap, there are two possible strategies. The first, which *Open*GALEN calls "untangling" [9], is to deconstruct the individual ontologies into disjoint non-overlapping sub-ontologies, link these, and then recombine and reconstruct the original ontologies using composite concepts from the sub-ontologies. The classifier can then be used to identify equivalent concepts and relations between concepts. This is essentially the technique used by GALEN-IN-USE in reconciling terminologies of surgical procedures from different countries. The second strategy – variants of which are used by Franconi [10], Swartout [11] and Farquhar [12] – is to identify candidate equivalencies between the two source ontologies, assert that they are equivalent by axioms in the joint ontology, and then use the classifier to identify inconsistencies. The untangling technique is particularly appropriate to bridging scales and granularities and is the subject of this paper.

BRIDGING LEVELS OF SCALE

Bridging scales is aided by recognition of recurring principles in the relationships between and within scales. A useful way to categorise principles is according to two dimensions:

- Fractal vs scale-specific principles – *i.e.* whether or not the same pattern repeats at multiple scales.
- Within scale vs bridging scale principles – *e.g.* whether the principle applies within any one scale or links two or more scales.

---

[7] Sometimes termed "restrictions", "slot constraints" or "criteria". Note that GRAIL is unusual in using "sanctions" rather than descriptors involving "only" (∀).

[8] In the contexts of ontology languages and description logics, the universal quantifier "∀" is best understood as "only" rather than "all".

[9] "has-class" in the current DAML+OIL/OWL notation

Examples of all four possible types are given in Figure 1:

	Fractal	**Non-Fractal**
Within scale	discrete structures are made of continuous substances, *e.g.* collagen-ligaments	atoms are bonded in chemical moieties.
Bridging scales	Multiples of discrete structures make up continuous substances; *e.g.* cells-tissues;	molecular action is the mechanism for clinical effect, *e.g.* proton pump inhibition is the mechanism for inhibiting acid secretion.

Figure 1: Classification of principles

To illustrate how these principles work in more detail, examples of the partitive links from the extended *Open*GAIEN ontology [6] are given in Figure 2. (In each case the relation given is only the parent relation of a hierarchy.)

In general, partonomic links are fractal: some hold within each scale, some linking pairs of scales at successive levels. The pattern by which *Multiples* of discrete *Structures* at one scale "make up" mass *Substances* at higher scales is particularly important. It holds for example from molecules to substances, cells to tissues, and even people to crowds, although it breaks down at the quantum scale. ("Multiples" are not sets – their identity is not derived strictly from their membership. A "bar of steel", "my liver tissue", and "the crowd at the fair" may all be considered to have a continuous identity even though all the individual units which make them up have been replaced. See Welty and Guarino [8])

Some functional links are similarly fractal – *Processes* "act on" *Substances*, *Structures*, and other *Processes*. However, many more functional links are specific to a given scale, *e.g.* *Genes*[10] "code for" *Proteins* and *Enzymes* "catalyse" *Reactions* at the molecular scale. The notion that enzyme actions "mediate" *Physiologic Response* links the molecular and higher micro and macro scales. Specific *Bindings* at the molecular scale are "the mechanism for" *Membrane transport* at the cellular scale, etc.

Scales are not absolute. They are, at least in part, a matter of our conceptualisations. Nature does not always oblige by providing entities to match those conceptualisations. A special problem arises when notions which are usually at two different scales appear at a single scale – *e.g.* for singular cellular organisms the organism and the cell coincide. In these cases it is usually best to maintain a consistent pattern even though the 'scales' are artificial, *e.g.* to treat a "bacterium" as being made up of a single "bacterial cell" and to speak of the "wall of the cell of the bacterium" rather than the "wall of the bacterium".

MANAGING CONTEXT: VARIANTS OF RELATIONSHIPS AND CONCEPTS

Meaning varies with context in normal language. In formal representation, context must be represented explicitly.

Handling context in traditional logic-based representation formalisms has been difficult because ontology languages developed prior to the mid 1990s (*e.g.* Classic, Loom, KRSS, etc.) had no constructs to make additional descriptive statements about defined concepts, so it was not possible to make statements about "All Human Hands", "All Anatomically normal Lungs", etc. One of GRAIL's important features was to provide such constructs in the form of "necessary statements", albeit the reasoner for them was not logically complete. One of the major contributions of Horrocks' development of the FaCT reasoner underpinning DAML+OIL/OWL[13] is that it provides logically complete tractable inferences with such constructs (now known as "general inclusion axioms"). Likewise, many early formalisms did not provide for a hierarchy of properties semantic links ("properties") but only of concepts. Again, GRAIL provided an initial partial solution, and Horrock's FaCT reasoner has provided logically complete inference.

Given support for both the further description of defined concepts and for a hierarchy of semantic links, there are two approaches to dealing with context:

- By including context markers in composite concepts, *e.g.* "'*anatomically normal* hands' have five fingers"
- By using variants of the semantic link to convey subtly

Fractal	**within / between**	**Scale**	**Object**	**Link**	**Target**	**Example**
yes	within	macro / micro / cell	Substance	makesUpInPart	Structure	Collagen-Tendon
			Structure	isComponentOf isSubdivisionOf isLayerOf contains	Structure	Valve-Heart Ventricle-Heart Myocardium-Heart Chest-Heart
			Substance	isPortionOf	Substance	Water-Blood
	between	macro / micro to cell	Substance	isMadeOfMultipleOf	MicroStructure	Tissue-Cell
no	between	macro / micro / cell to chemical	Substance	isMadeOfChemically	ChemicalStructure	Water - WaterMolecule
	within	chemical	ChemicalStructure	hasBondedIn	ChemicalStructure	WaterMolecule-OxygenAtom

Figure 2: Example set of Fractal and Scale-specfic, within and between scale semantic links for partitive relations

different meanings, *e.g.* to convey the strict structural view of part-whole in the Foundational Model of Anatomy separately from a functional and clinical view of part-whole relations needed for clinical applications.

Including context markers in composite concepts

The first approach – including context markers in composite concepts – is appropriate where there are numerous subtly different concepts which have much in common but where some things depend on context. One example is in dealing with anatomically normal and variant structures. For example, we can indicate that, in all possible contexts, all hands are subdivisions of the upper extremity, but add that in the "*AnatomicallyNormal*" context, hands have exactly five fingers:

Hand → isSubdivisionOf some UpperExtremity

Hand & AnatomicallyNormal →
hasSubdivision exactly-5 fingers

In a language with a complete reasoner such as DAML+OIL/OWL, the reasoner will infer that any hand with fewer or more fingers does not fit the context *AnatomicallyNormal*. In GRAIL which lacks negation, and whose reasoner is not complete, additional axioms are necessary to express this inference directly, *e.g*:

Hand & hasSubdivision Finger6 →
AnatomicallyNonNormal

In either case, the reasoner can determine whether or not a given concept fits a given context, but the DAML+OIL/OWL solution is clearly more direct although more computationally expensive.

These same mechanisms can be used to express contexts based on differences between species, homologies or orthologies. For example, there are many facts about hands that are common in the context "Primate", *e.g.*

Hand of Primate → hasSubdivision exactly-1 Thumb

but some which are different in the subcontext "Human" and "NonHumanPrimate" *e.g.*:

Thumb of Hand of Human →
hasFeature Opposable
Thumb of Hand of NonHumanPrimate →
not hasFeature Opposable

Using variant semantic links to convey context

The second mechanism – using the hierarchy of semantic links – is appropriate where there are subtly different views about the meaning of relations, as between the Digital Anatomist Foundational Model of Anatomy strict structural view of partonomy and the looser view taken by clinicians. The technique is illustrated in detail in Figure 3. Structurally, the pericardium is a separate organ from the heart as represented in the Foundational Model of Anatomy. However, clinically, we wish "pericardial disease" to be classified with "diseases of the heart" just as other diseases of parts of the heart are classified as diseases of the heart.

```
Link Hierarchy

isGeneralPartOf
  isStructuralPartOf
    isFMAStructuralPartOf
      isSubdivisionOf
      isComponentOf
    ...
  isFunctionalPartOf
  isClinicalPartOf

Resulting hierarchy of composite concepts
BodyPart wh ich isGeneralPartOf some Heart
   BodyPart which isFMAStructuralPartOf some Heart
     CardiacAtrium
     ...
  BodyPart which isClinicallyPartOf some Heart
    Pericardium

Clinical definition of Heart Disease and Pericarditis
  HeartDisease ≡
    Disease which hasLocation
      (some Heart or
        (BodyPart which isGenerallyPartOf some Heart))

  Pericarditis ≡
     Inflammation which hasLocation some Pericardium

Resulting classification of Pericarditis under Heart Disease
  HeartDisease
   Pericarditis
```

Figure 3: Extract of semantic link hierarchy for part-whole relations showing distinctions between functional, structural, and strict structural for Foundational Model of Anatomy (FMA), plus example of achieving classification of Pericarditis as a heart disease without disturbing the FMA classification.

To achieve this, as shown in Figure 3, we use two sibling partonomic relations "*isStructurallyPartOf*" and "*isClinicallyPartOf*", with a common parent "*isGenerallyPartOf*".

Using this construction it is possible to query the knowledge base either for the precise:

BodyPart & isStructuralPartOf some Heart

(ie not including the pericardium), or for the more general

BodyPart & isGeneralPartOf some Heart

Since being either a *ClinicalPart* or a *StructuralPart* implies being a *GeneralPart*, this allows us to add variant notions of part-whole relations without corrupting the careful design of the Digital Anatomist's stricter notion of partonomy.

If we could be certain that being a "structural part" always implied being a "clinical part", then *isClinicalPartOf* could subsume *isStructuralPartOf* directly. However, although no one has yet suggested a counter example, it is methodologically cleaner in early experimentation to make each potential flavour of context, such as partonomy, clearly separate and provide a common parent. For the same reasons we have distinguished *isFMAStructuralPartOf* from *isStructuralPartOf* to allow for the possibility that these are alternative views of structural partonomy specifically, even though we expect to show them to be equivalent.

DISCUSSION

Issues of scale and context pervade knowledge representation. The experience of *Open*GALEN and related projects on the UK Drug Ontology [14] is that notions of scale tend to emerge from the process of normalising and "untangling" taxonomies whereas notions of context tend to emerge from the conflicts between disciplines.

Notions of scale are often cited as the major problems, but in our experience they follow a comparatively regular and therefore tractable pattern, provided authors are careful to distinguish notions with similar names at different scales – *e.g.* "water" from "water molecule".

The key requirement is that the ontologies at each scale be "normalised" and disjoint. Given this assumption, formal methods can guarantee correct classification of the resulting linked ontology. Correspondingly, errors in the classification often indicate violations of that assumption, typically through inclusion of concepts with overlapping, or ambiguous, interpretation. Systematic examination of the classification for specific types of errors is therefore an important part of the empirical validation of that the assumption holds. Since the meanings of concepts are ultimately a matter of human interpretation, the validity of the assumption must always be tested empirically.

More work remains to be done on managing context, but the tools provided by expressive ontology languages such as DAML+OIL/OWL allow experiments to be conducted flexibly, using the reasoner underlying the language to identify the consequences of the experiments. Such languages are expressive enough to support a methodology which defers 'ontological commitment' until empirical evidence is available to support it – *e.g.* whether or not the clinical notion of partonomy always subsumes the structural.

An important caveat is that none of these mechanisms, nor any others likely to solve these problems, are intuitive to general users. The logic-based ontology employing these mechanisms is best regarded as an "assembly language" to be hidden from users by more intuitive domain-oriented "Intermediate Representations" as described elsewhere [4][9].

Initial experiments with the Digital Anatomist Foundational Model of Anatomy (FMA), with mouse anatomy and with congenital abnormalities all indicate the flexibility of the approach. However, experiments to date are of modest scale, and it cannot yet be said with certainty that the methods scale up to ontologies with tens or hundreds of thousands of concepts. Likewise, there remain questions about how much complexity can be combined in a single ontology without encountering insurmountable computational obstacles.

Despite these reservations, current experience is sufficient to indicate the usefulness and wide applicability of the techniques. The fundamentally fractal nature of formal ontologies, their ability to combine independent subontologies, and their ability to manage subtle differences in the meaning of semantic links makes them important tools for the bridging of scales and contexts typical of bioinformatics and clinical informatics.

References

1. Gruber TR. A translation approach to portable ontologies. *Knowledge Acquisition* 1993;5(2):199-220.
2. Uschold M, Gruninger M. Ontologies: principles, methods and applications. *Knowledge Eng Rev* 1996;11(2).
3. Spackman KA, Campbell KE, Côté RA. SNOMED-RT: A reference Terminology for Health Care. *J Am Med Inf Assoc (JAMIA)* 1997(Symposium special issue):640-644.
4. Rector AL. Clinical Terminology: Why is it so hard? *Meth Inf Med* 1999;38:239-252.
5. Fensel D, van Harmelen F, Horrocks I, McGuinness D, Patel-Schneider P. OIL: An ontology infrastructure for the semantic web. *IEEE Intelligent Systems* 2001;16(2):38-45.
6. Rector A, Rogers J. Ontological issues in using a description logic to represent medical concepts: Experience from GALEN. *Meth Inf Med* 2002(in press).
7. Rosse C, Shapiro IG, Brinkley JF. The Digital Anatomist foundational model: Principles for defining and structuring its concept domain. *J Am Med Inf Assoc* 1998(1998 Fall Symposium Special issue):820-824.
8. Welty C, Guarino N. Supporting ontological analysis of taxonomic relationships. Data and Knowledge Engineering 2001.
9. Rector A, Wroe C, Rogers J, Roberts A. Untangling taxonomies and relationships: Personal and practical problems in loosely coupled development of large ontologies. In: Gil Y, Musen M, Shavlik J, editors. *Proc First Int Conf on Knowledge Capture (K-CAP 2001);* ; Victoria, BC, Canada: ACM; 2001. p. 139-146.
10. Franconi E. A Semantic approach for schema evolution and versioning in object-oriented databases. In: *6th IntConf on Rules and Objects in Databases (DOOD'2000);* 2000;
11. Swartout W, Patil R, Knight K, Russ T. Toward distributed use of large-scale ontologies. In: Gaines B, Musen M, eds *Tenth Annual Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW96);* ; Banf,Canada; 1996.
12. Farquhar A, Fikes R, Pratt W, Rice J. Collaborative Ontology Construction for Information Integration. Technical Report: Knowledge Systems Laboratory, Stanford University; 1995. Report No.: KSL-95-63.
13. Horrocks I, Staller U, Tobies S. Practical reasoning for very expressive description logics. *Journal of the Interest Group in Pure and Applied Logics (IGPL)* 2000;8(3):293-323.
14. Solomon WD, Wroe C, Rogers JE, Rector A. A reference terminology for drugs. *J Am Med Inf Assoc* 1999(Fall Symposium Special Issue):152-155.