

# Granularity Scale & Collectivity: When size does and doesn't matter

Alan Rector & Jeremy Rogers  
Department of Computer Science  
University of Manchester  
Manchester M13 9PL, UK  
rector@cs.man.ac.uk

**Abstract:** Bridging levels of “granularity” and “scale” are frequently cited as key problems for biomedical informatics. However, detailed accounts of what is meant by these terms are sparse in the literature. We argue for distinguishing two notions: “size range”, which deals with physical size, and “collectivity”, which deals with aggregations of individuals into collections which have emergent properties and effects. We further distinguish these notions from “specialisation”, “degree of detail”, “density” and “connectivity.” We argue that the notion of “collectivity” – molecules in water, cells in tissues, people in crowds, stars in galaxies – has been neglected but is a key to representing biological notions, that it is a pervasive notion across size ranges – micro, macro, cosmological, etc – and that it provides an account of a number of troublesome issues including the most important cases of when the part-whole relation is, or is not, transitive. Although examples are taken from biomedicine, we believe these notions to have wider application.

## 1. Introduction

It is a truism that a major challenge for bioinformatics is to bridge levels of granularity and scale, from molecular, to cellular, to organ, to organism, to ecology. However, it is rarely made clear exactly what is meant by “granularity” or “scale” or what the consequences are of differences in granularity and scale for which any explanation must account.

This paper argues that it would be clearer to distinguish unambiguously two dimensions. We term these two dimensions “collectivity” and “size range” despite the risk of adding yet further neologisms to the field<sup>1</sup>. The basic notion that we put forward is that entities considered individually at one level are considered as collectives at the next level – *e.g.* collectives grains of sand form a beach, collectives of stars form galaxies, collectives of cells form tissues. In general, for convenience, we shall refer to the “grains” of a “collective” and correspondingly to “granular parts”. The notion of “collective” used here is similar to that of “groups” used by Artale [1, 2] and by Winston & Odell [11, 28], but neither they nor Padgham & Lambrix [12] investigate it extensively, although Winston and Odell put forward analogous reasoning for why the feet of geese do not form part of a flock. The notion of “granular parts” is also hinted at by the distinction between “constituent parts” and other forms of part-whole relation in the Foundational Model of Anatomy [18]. However, we suggest that this is a seriously under investigated aspect of representation and can be used to account for several important phenomena.

Our fundamental contention is that there are properties and effects of collectives that are emergent – *i.e.* that are not generally predictable from the properties of the individual grains and that therefore must be attributed to the collective as a whole: the mood of a crowd is distinct from that of its constituent individuals; a beach has area; galaxies have mass; tissues have structure, strength, and are responsible for many effects such as secretion or absorption. The features and effects that are mentioned are of the collectives rather than the individual grains that make them up. More specifically we seek:

- To distinguish the way in which, for example, a cell is part of the body from the way a finger is part of the body – specifically that the loss of a cell not diminish the body whereas the loss of a finger does;
- To use this to motivate an important criteria for when transitive relations should, or should not, be treated as transitive;
- To represent loosely repetitive patterns in tissues – that the “cells in the mucosa are aligned”;

---

<sup>1</sup> Although we would prefer to reserve the term “granularity” for the notion here termed “collectivity”, the term “granularity” has become so overloaded with different meanings in different fields that we reluctantly opt for a neologism rather than risk further confusion and controversy. “Scale” conforms more closely to “size”. However, to avoid confusion we have likewise been explicit in this paper and used the term “size range”.

- To deal with the collective effects of cells, organelles, etc. – *e.g.* the process of secretion and regulation of hormones by the cells of endocrine organs or the collective strength of muscles made up of indeterminate numbers of muscle fibres.

More often than not, collectives are themselves portions of larger things. Galaxies are more than mere collectives of stars; tissues are certainly more than collectives of cells; even a beach is more than a collective of sand.

However in each case there is a sense in which we can treat the collective as a portion of the larger entity. If we can have independently measurable commensurable features for both the collective and the larger entity, we can speak of the proportion of the greater entity formed by the collective of grains just as we would speak of the proportion of water or salt in an amount of sea water, collagen in tissue, or the proportion of the mass of galaxy comprised of the visible stars.

Our goal is a set of broadly applicable principles. The paper follows broadly the intent and lessons, although not always the execution, of the *OpenGALEN* Common Reference Model[14, 17]. As an illustration we present this paper and an implementation in the framework of OWL-DL<sup>2</sup>. However, the issues are general and independent of any particular implementation. An OWL-DL ontology illustrating the principles can be found at <http://www.cs.man.ac.uk/~rector/ontologies/collectivity-demo.owl> and <http://www.cs.man.ac.uk/~rector/ontologies/collectivity-demo-classified.owl>

## 1.1 Outline of approach

We distinguish two notions often confused under the heading of “granularity”:

**Collectivity** – *Grains vs Collectives* – the degree of collectivisation, *e.g.* with respect to water filling a lake, the relation ‘filling’ is to the water as, amongst other things, a collective of water molecules, not to the individual molecules themselves.

**Size range**– *Large vs Small* – the size of an object with respect to the phenomena that affect it, *e.g.* quantum scales of distance or relativistic scales of speed. However, less extreme differences in scale can have major effects. Surface tension is critical at the scale of a water flea’s interaction with water but not at that for a human.

Our major contentions are that:

- “Collectivity” and “size range” are distinct notions.
- Two patterns and subrelations of the basic part-whole relations should be distinguished, one for “determinate parts” such as finger is to hand, and one for “granular” parts such as the cells of the tissues that make up the finger. Determinate parts are directly part of the whole and of a determinate number: removing one damages or diminishes the whole. Granular parts are parts by way of being members of a collective that is part of the whole and of indeterminate in number: removing one does not (normally) diminish the whole.
- The part-whole relation for determinate parts is transitive, and the part-whole relation for granular parts is not. More specifically, the membership relation between individual grains and collectives is not transitive although it is a subproperty of the generic part-whole relation.
- There are effects and characteristics concerned with the emergent characteristics of collectives as a whole not determinable from the characteristics of their grains, even where those characteristics can be determined for the grains individually. Such effects and characteristics should be represented as pertaining to the collective rather than to its grains.

Our basic distinction between granular and determinate parts is whether there is a fixed, or nearly fixed number of parts – *e.g.* in fingers of the hand, chambers of the heart, or wheels on a car, eggs in a follicle – or an indeterminate number referred to merely by the collective – cells in the skin of the hand, red cells in blood, or rubber molecules in the tread of the tire. We accept that there are borderline cases, and that we sometimes wish to refer to the collective properties of a fixed number of entities. We will return to this issue towards the end of this paper after the basic notions are established. (See Section 4.3.1.)

A summary of the basic hierarchy of properties in OWL that we propose to implement these notions is given in Table 1. The elisions will be explained later in the paper.

is_part_of / has_part	transitive
— ...	
— is_determined_part_of / has_determined_part	transitive
— —...	
— is_grain_of / has_grain	not transitive

<sup>2</sup> An OWL-DL ontologies illustrating the principles can be found at <http://www.cs.man.ac.uk/~rector/ontologies/collectivity>

**Table 1: Simplification of hierarchy of partonomy relations and their inverses.**

## 1.2 Other notions sometimes labelled “granularity”

We further distinguish “collectivity” and “size range” from four other notions with which they may be confused, and which other researchers have referred to as ‘granularity’ in addressing mereological issues [4, 22].

**Specialisation** – *Category vs kind*– the usual notion of “is-kind-of”, *e.g.* that “mammal” is a generalisation including, amongst other things, dogs and elephants. Sometimes also labelled ‘abstraction’.

**Degree of detail** – The amount of information represented about each entity, regardless of its level of specialisation. Crudely in an ontology represented in OWL, the number of axioms and restrictions concerning each entity.

**Density** – The number of semantically ‘similar’ concepts in a particular conceptual region. How “bushy” the subsumption graph is. High local density in an ontology usually co-occurs with high levels of specialisation and degree of detail, but in two different ontologies of the same overall depth, in a particular section one may find the same two categories separated by different numbers of intervening categories or possessing very different numbers of sibling categories.

**Connectivity**– The number of entities connected directly and indirectly to a given entity either through generalisation/specialisation or by other properties.

These notions will not be further discussed in this paper.

## 1.3 Criteria for success of the proposed approach

Our purpose in developing “ontologies” is to support information systems. The test of their adequacy is whether they can effectively represent the entities about which information must be communicated so that that communication is “faithful”. This focuses our interest as much on the relations (“properties” in OWL; “roles” in most DLs; “attributes” in GRAIL) as on the entities related.

Our specific application is biomedicine, so that we will test our solution primarily with respect to well known biomedical knowledge resources including the Digital Anatomist Foundational Model of Anatomy [10, 18], the Open Biology Ontology (OBO) and more particularly the Gene Ontology [3, 26, 29] and *OpenGALEN* [15-17]. Secondly Johansson’s paper provides a series of examples against which to test the notions put forward here {{ref}}.

More specifically, we seek a set of patterns and relations in OWL that are adequate to capture four notions and exclude their counterexamples:

- Relation of faults and procedures to parts and wholes – *e.g.* that the disease of the part is a disease of the whole and certain procedure – *e.g.* repair – on a part is a procedures on the whole.
- Patterns and characteristics of collectives *e.g.* that the cells of the intestine are typically aligned (with each other) or that the cells in bone are sparsely separated.
- Collective or emergent effects of collectives, *e.g.* the total secretion of enzymes by the liver cells or the total force exerted by the cells in a muscle.
- Persistent vs non-persistent parthood – *e.g.* that “Jack’s finger” will still be referred to as “Jack’s finger” even when it is severed from his hand. However, insulin secreted by a cell is not considered to be a part of that cell.

## 1.4 Notation

Neither of the XML concrete syntaxes for OWL is compact or readable enough for easy use in a paper, and even the official abstract syntax becomes bulky and difficult to read when there is any significant embedding. This paper therefore adopts the following conventions for a simplified syntax. Complete working OWL DL implementations can be found at the link indicated.

- Subset and subproperties are indicated by indentation made explicit by ‘-’s. Where only two are involved a simple arrow is used, *e.g.* “Heart → Organ” for “Heart is a kind of Organ”.
- Properties are presented with their inverse separated by a slash, property modifiers – transitive, symmetric, functional, etc. are listed to the right, as in Table 1 above.
- The OWL key words are adapted to a concise infix notation as shown in Table 2.
- In complex expressions, indentation will be used rather than bracketing wherever the meaning is clear
- Schema variables will be given in italics sans serif in place of parts of names, *e.g.* *X,Y,Z* as in *part\_of X*. Schema variables range over OWL class names.

Abbreviated Informal	OWL Abstract Syntax	DL German Syntax
A AND B	intersectionOf(A B)	$A \sqcap B$
A OR B	unionOf(A B)	$A \sqcup B$
NOT A	complementOf(A)	$\neg A$
has_property SOME C	restriction(has_property someValuesFrom(C))	$\exists \text{ has\_property . } C$
has_property ONLY C	restriction(has_property allValuesFrom(C))	$\forall \text{ has\_property . } C$
$A \rightarrow B$	subclassOf(A B)	$A \sqsubseteq B$
$A \equiv B$	equivalentClass(A B)	$A \equiv B$

**Table 2: Concise infix notation used in this paper with equivalents in OWL and standard DL notation**

## 2. Basic Notions: Collectivity and Size

The notion of “collectivity” that is proposed here turns on whether the information to be conveyed relate to individuals or to “collectives” of individuals. “Collectives” have several properties in common:

- They are not mathematical sets, because their identity does not depend on exact membership, *i.e.* on their extension; *e.g.* the grains of sand on the beach may be completely changed and yet we call it the same beach.
- They persist as parts even when individual grains are shed or gained – *e.g.* the skin persists as part of the body even though its cells are constantly shed and renewed.
- They have emergent properties that do not pertain to individuals, *e.g.* cells may be aligned or distributed, tiles in a wall may create a pattern, etc. The alignment or pattern is not a property of the individual grains but of the collective and furthermore of the collective to the whole which it constitutes.
- They have emergent effects that cannot be attributed to their individual grains, *e.g.* the property of water that it expands when frozen emerges from the properties of water molecules collectively, but cannot be attributed to individual molecule.

### 2.1 “Collectivity” does not depend on physical size

Necessarily, grains are not physically larger than the collective of which they are members (except perhaps for some odd quantum cases). There is a tendency to talk of things as being at, for example, the “cellular level” or the “organ level” or the “subatomic” level, etc. However, such talk indicates a general tendency and conflates size and collectivity. Hairs are macroscopic entities of the same general size as small organs, yet most of the information we have to convey about hairs concerns the collective “hair” rather than individual “hairs”. Sperm and eggs are both cells, but much of what we have to say about eggs pertains to individual eggs, whereas much more that we have to convey about sperm concern the collective, although we need a mechanism to cross levels of collectivity to speak of a single sperm fertilizing a single egg. Indeed, one of the issues in fertility research is to determine which factors depend on the collective of sperm and the fluids in which they are swimming, and which depend on the individual sperm cells themselves.

To extend the biological examples, within the cells there are both individual entities, such as the nucleus, and collectives such as mitochondria and chloroplasts. Within the nucleus there are a precisely countable number of chromosomes, which are usually treated individually, but uncountable collectives of macromolecules. In some circumstances, the same entities at the same size may be sometimes treated collectively and sometimes individually. The rigidity and shape of a chromosome are a collective property of the DNA molecules (and other supporting structures) that make it up; the “genes”<sup>3</sup> inheritance of characteristics is usually a feature of discrete sequences of base pairs (with complex dependence on context and regulation).

### 2.2 “Size range” depends on physical dimension but not on granularity

There are many effects that are specific to physical size, distance, speed, density, etc. Most obviously, quantum and relativistic effects are generally relevant only for the very small, very large or the very rapidly moving<sup>4</sup>. Closer to everyday life, the surface tension and vortex effects that govern insects ability to fly, walk on walls, skim over water, etc. are highly relevant at their size range but almost irrelevant at the size of most mammals. Within biology, chemical bonding, van der Waals forces, other electrostatic forces, and many other effects are important at one physical size range but not at another. When they are relevant, they are relevant for both for individuals and collectives that conform to that size range.

<sup>3</sup> The definition of what constitutes a gene is problematic, at least in eukaryotic cells, but that need not concern us here.

<sup>4</sup> relative to the observer of course.

### 3. Semi Formal Presentation

#### 3.1 Basic properties and entities

We shall assume an upper ontology similar to DOLCE that includes the notions of “Physical material entity” termed “Physical\_object”. We shall assume a parent “Physical\_entity” that includes both material entities and non-material entities such as holes and lines. We shall assume a notion of “Amount of matter” as in DOLCE but leave open until later the discussion of the controversy between cognitivist and realist over the nature of the link between amounts of matter and physical objects. Correspondingly we shall leave open at this point the elaboration of the structure under *is\_ingredient\_of*. We shall assume a basic part-whole relation conforming in principle to the axioms of mereology (although these cannot be fully captured in OWL), termed “*is\_part\_of / has\_part*”.

Using the conventions described in 1.4 above we then present the basic property hierarchy in table 3a.

Property	Transitive	Domain/ Range	Comments
<i>is_part_of / has_part</i>	Y	Physical_entity / Physical_entity	The generic part-whole relation Reflexive & antisymmetric properties not captured directly in OWL.
— <i>is_gross_part_of / has_gross_part</i>	Y	Physical_entity / Physical_entity	The common parent of measurable portions and determinate parts.
— — <i>is_determinate_part_of / has_determinate_part</i>	Y	Physical_entity / Physical_entity	The relation between determinate parts and wholes, <i>e.g.</i> fingers and hands.
— — <i>is_ingredient_of / has_ingredient</i>	Y	Amount_of_matter/ Amount_of_matter	Further details deferred – see 3.2.3
— — ...			See Section 3.3
— <i>is_grain_of / has_grain</i>	N	Physical_object / Collective	The relation between a grain and the collective. (NB: The status of collectives of non-material entities is left open at this point)

**Table 3a**

The corresponding entity hierarchy is described in table 3b.

Class	Use in this paper	Comments
Physical_entity	Domain/range of <i>is_part_of</i> and <i>is_determinate_part_of</i>	Common ancestor of all physical entities
— Physical_object	Domain for <i>is_grain_of</i>	Material physical entities
— Non_material_object	Excluded from domain for <i>is_grain_of</i>	Non-material physical entities, <i>e.g.</i> holes, lines, etc.
— Amount_of_matter	range for <i>is_ingredient_of</i>	Amounts of “stuff”, roughly corresponding to mass nouns. (NB the Relation between <i>Physical_object</i> and <i>Amount_of_matter</i> depends on the debate between the cognitivist & realist stance and is not directly relevant to this paper. See 3.3)
— — Mixture	domain for <i>is_ingredient_of</i>	Abstract including solutions, suspensions etc.
Collective	Range of <i>is_grain_of</i>	Whether or not Collectives are considered physical and whether or not they are to be disjoint from <i>Physical_object</i> , is deferred. See 4.3.5

**Table 3b**

## 3.2 Basic schemas

### 3.2.1 Defining collectives

Normally collectives are defined using universal restrictions, *i.e.* collectives are defined following a schema where the upper case *Italics* indicates schema variables that range of class names.

$Collective\_of\_X \equiv Collective \text{ AND } has\_grain \text{ ONLY } X$

There are two consequences of this schema:

- Empty collectives are allowed. This is convenient when we want to talk about concentrations of zero or things that are empty or missing. We can define `Non_empty_collective` in the obvious way as `Collective AND has_grain SOME Anything`<sup>5</sup>
- A collective must be defined in terms of the most general type of which its grains are individuals, which may be a disjunction. (However, any collective defined in terms of a disjunction should be viewed with suspicion, as it is more likely to be more appropriately represented as a mixture.)

### 3.2.2 Reflexive parts

Because reflexive properties cannot be expressed directly in OWL it is useful to define two schemas for reflexive parthood:

$Reflexive\_part\_of\_X \equiv X \text{ OR } is\_part\_of \text{ SOME } X$   
 $Reflexive\_gross\_part\_of\_X \equiv X \text{ OR } is\_gross\_part\_of \text{ SOME } X$   
 $Reflexive\_determinate\_part\_of\_X \equiv X \text{ OR } is\_determinate\_part\_of \text{ SOME } X$

Which schema is appropriate depends on the requirement. In simple “part explosions” only determinate parts are required. If both portions and determinate parts are required (see “Mixtures” below), then `Reflexive_gross_part_of_X` is required. If all parts, as in the digital anatomist or classic mereology, the most general notion of `Reflexive_part_of_X` is required.

In effect, these schemas allow much of the effect of reflexivity to be transferred to the entities from properties. It also makes it easy to express the notions related to SEP Triples [7, 8, 19]. Which is used in which circumstances depends on whether the class to be defined is to include only parts within the given level of collectivity or whether it is to include parts across levels of collectivity.

### 3.2.3 Mixtures

Collectives and reflexive parts provide the basic mechanisms required, but almost all interesting cases involving collectives involve not just one collective but mixtures of collectives with other collectives and/or amounts of matter.

The relation between different collectives and amounts of matter in a mixture we term `is_ingredient_of / has_ingredient`, which is transitive. We place `is_ingredient_of` as a sibling of `is_determinate_part_of` and under `is_gross_part_of` because common classes and queries to be formulated include both, *e.g.* the gross parts of a car include both wheels and rubber; the gross parts of the arm include both the biceps and fascia.

The basic schema for mixtures is:

$Mixture\_of\_X_1\_and\_X_2\_and\_...\_and\_X_n \equiv$   
 $Mixture \text{ AND } has\_ingredient \text{ SOME } X_1 \text{ AND } has\_ingredient \text{ SOME } X_2 \text{ AND } \dots \text{ AND } has\_ingredient \text{ SOME } X_n$

Formally, the domain constraint on `is_ingredient_of` guarantees in this simple version that anything that has portions is a mixture. However, for clarity it is better to include `Mixture` as a conjunct explicitly. A `Mixture` can be defined by being an amount of matter that has ingredients<sup>6</sup>.

$Mixture \equiv Amount\_of\_matter \text{ AND } has\_ingredient \text{ SOME } Amount\_of\_matter$

For example, one might represent that blood is a mixture of – amongst other things – plasma, red cells and white cells:

$Amount\_of\_blood \rightarrow$   
 $Mixture \text{ AND }$   
 $has\_ingredient \text{ SOME } Plasma \text{ AND }$   
 $has\_ingredient \text{ SOME } (Collective \text{ AND } has\_grain \text{ ONLY } White\_blood\_cell) \text{ AND }$   
 $has\_ingredient \text{ SOME } (Collective \text{ AND } has\_grain \text{ ONLY } Red\_blood\_cell)$

Note that, in common with most biomedical definitions, we have not closed the list of ingredients in the mixture. There is nothing in the above axiom to imply that blood does not contain other things, only that it does contain the

---

<sup>5</sup> owl:Thing

<sup>6</sup> A given ontology might, for consistency, wish to insist that all amounts of matter were mixtures. That issue is deferred here

ingredients mentioned. Nor have we made this a definition; it does not imply that any mixture of plasma, red cells and white cells is blood.

### 3.2.4 Proportions

Because the relative amounts in a mixture are so often important, and because the means of determining relative amounts vary – *e.g.* by weight, volume, activity, etc.– in a binary relational formalisms such as RDF or OWL, it is often appropriate to reify the relation `has_ingredient`, *i.e.* to re-represent it as a class, which we shall term Proportion.

The basic schema is:

```
Mixture_of_X1_and_X2_and_..._and_Xn ≡
  Mixture AND
  has_proportion SOME (Proportion AND is_of_ingredient SOME X1 AND has_percentage VALUE p1) AND
  has_proportion SOME (Proportion AND is_of_ingredient SOME X2 AND has_percentage VALUE p2) AND
  ...AND
  has_proportion SOME (Proportion AND is_of_ingredient SOME Xn AND has_percentage VALUE pn)
```

The example of blood extended to this schema becomes::

```
Amount_of_blood →:
  Mixture AND
  has_proportion SOME (Proportion AND is_of_ingredient SOME Plasma
    AND has_percentage VALUE p1) AND
  has_proportion SOME (Proportion AND is_of_ingredient (Collective AND has_grain ONLY White_blood_cell)
    AND has_percentage VALUE p2) AND
  has_proportion SOME (Proportion AND is_of_ingredient (Collective AND has_grain ONLY Red_blood_cell)
    AND has_percentage VALUE p3)
```

In principle the  $p_i$  can be either numeric values or ranges, precise or fuzzy. In practice as OWL exists today, pending technical issues around user defined XML datatypes and issues with classifiers, they are confined to integers. Additional qualifiers on Proportion include the mechanism – *e.g.* by mass, volume or some other measure.

### 3.2.5 Allowing proportions and simple ingredients to coexist

It is possible to allow the two patterns – for simple ingredients and for proportions of ingredients to coexist if we arrange the property hierarchy as shown in Table 4. Care must be taken with the domain and range constraints.

Property	Transitive	Domain/ Range	Comments
<code>is_ingredient_of /</code> <b><code>has_ingredient</code></b>	Y	Amount_of_matter / Amount_of_matter OR Proportion_of_matter	Ingredients of ingredients are ingredients of the whole
<code>— of_mixture /</code> <b><code>has_proportion</code></b>	N	Proportion / Amount_of_matter	Proportions of proportions are not proportions of the whole.
<code>— is_proportion /</code> <b><code>is_of_ingredient</code></b>	N	Amount_of_matter/ Proportion	

**Table 4: Property hierarchy reconciling ingredients and proportions. Note that the relevant properties are the inverses (given in bold) to remain consistent with Table 3a.**

The fact that proportions of proportions are not themselves the same proportions of the whole is reflected in the facts that `has_proportion` and `is_of_ingredient` are not transitive. Since the percentages attached to each proportion will have to be recalculated at each step down the chain, the relationship is not simply transitive but follows a more complex rule. That rule must be handled by reasoning mechanisms outside the scope of OWL or most other ontology languages. What can be captured in OWL is that ingredients of ingredients, by either mechanism, are ingredients of the whole, which is represented by the fact that the parent property, `has_ingredient`, is transitive.

### 3.2.6 Characteristics of collectives and patterns of collectives in mixtures

*Characteristics of the collective itself.* Members of a collective often have collective characteristics, *e.g.* that the cells of a tissue are aligned or that the atoms of a crystal form a particular lattice structure, that neurons fire synchronously or asynchronously, etc. Such characteristics pertain to the collective; they make no sense if applied to its individual grains. Nor do these characteristics depend on the collective's relation to any other entity of which it may be a part. Furthermore, just as collective's identity is not extensional, their characteristics are not universal over their extensions, *i.e.* they can be considered true even if they do not apply to every member of the collective, *e.g.* a crystal will still be said to have a particular alignment even if it has flaws.<sup>7</sup> Hence it is appropriate to represent such characteristics as properties of the collective, *e.g.*

<sup>7</sup> How completely such characteristics are true belongs with a discussion of fuzziness or precision and is beyond the scope of this paper.



“‘Amount of clay’ *constitutes* the ‘Statue’”. Smith’s ‘Realist/Unitary’ view is that there is a single entity and that the “‘Amount\_of\_clay’ *is* the ‘Statue’”, or more precisely that the “‘Amount\_of\_clay’ *is* (during some time span) the ‘Statue’”.

The issues discussed in this paper are largely independent of this controversy. For purposes of this paper and presentation in OWL, the factorisation provided by the Cognitivist/Multiplicative view is clearer and briefer, so we shall adopt it here. To do so requires adding the relation *constitutes/is\_constituted\_by* to table 3a at the point marked by the elision as one of the additional kinds of “gross parthood”. The domain of *constitutes* is *Physical\_object*, and the range is *Amount\_of\_matter*. Since the domain and range are different, and in most formulations disjoint, *constitutes/is\_constituted\_by* is non-transitive.

### 3.4 Use and consequences

#### 3.4.1 Propagation of faults

In general, faults propagate only across gross parthood, *e.g.* disorder to the liver is usually considered as a disorder of the digestive system, body, etc. whereas we would not normally consider a disorder of a single liver cell in this way. The liver cell is a grain of a collective that forms part of the liver (whether or not via a *constitutes* relation). Likewise, while we would consider a disorder of the metabolism of all, or a significant portion of, red cells – *e.g.* sickle cell anaemia – as a disorder of blood, we would not consider a disorder of the metabolism of a single red cell as a disorder of blood. Indeed, since both liver and red blood cells constantly die and are replenished, were we to consider the state of individual cells, all organisms would suffer from liver and blood disorders, which is clearly nonsense.

Hence the schema for disorders is normally

$\text{Disorder\_of\_X} \equiv \text{Disorder has\_locus SOME Reflexive\_gross\_part\_of\_X.}$

Where *has\_locus* is the property linking disorders to their anatomical or functional “site”. This captures the above two examples and analogous cases while excluding the case of damage to individual cells, etc.

Note that the issue of propagation across boundaries of collectivity is orthogonal to the issue of whether the disorder applies to the entity as a whole or to its reflexive parts. There are disorders – gastritis, inflammatory bowel disease, septicaemia (infection of the blood), etc. that refer to the whole taken as a whole rather than its parts. For these cases, the appropriate schema excludes all parts, whether gross or granular:

$\text{Disorder\_of\_X\_as\_a\_whole} \equiv \text{Disorder has\_locus X.}$

Furthermore, the issue is not dependent on size. Analogies can be found at all physical size ranges.

#### 3.4.2 Transitivity of part-whole relations

The issue of propagation of faults is closely related to the issue of transitivity of part-whole relations. Effectively, the argument in this paper is that most cases where the part-whole relation is not transitive involve transitions across levels of collectivity, *i.e.* involve chains of reasoning including the *is\_grain\_of* relation which is not transitive. Confusion arises because our usual language does not distinguish the broader *is\_part\_of* relation from its more specialised subrelations, here termed *is\_gross\_part\_of* and *is\_grain\_of*. The *is\_grain\_of* relation marks boundaries between levels of collectivity, or what are often called levels of granularity. However, we argue that the critical issue of whether the part-whole relation is transitive is not one of physical size, *per se*, but of whether or not the relation deals with collectives or individuals.

As a partial validation of this view, consider the list of cases provided by Johansson of anomalies where the part-whole relation is not considered to be transitive. Table 5 lists these issues and whether or not they are accounted for by the distinction between gross parthood and collective parthood.

1.	A handle, x, can be part of a door, y, and a door can be part of a house, z, but yet the handle need not be (is not) a part of the house. That is, ‘x < y’ and ‘y < z’ but ‘-(x < z)’. (Of course, ‘part’ cannot here and elsewhere in the list be synonymous with ‘spatial part’.)	<i>Not accounted for:</i> confusion of direct and indirect partonomy.
2.	A platoon is part of a company, and a company is part of a battalion, but yet a platoon is not part of a battalion.	<i>Not accounted for:</i> confusion of direct and indirect partonomy
3.	A cell’s nucleus is part of a cell, and a cell is part of an organ, but yet the nucleus is not part of an organ.	<i>Accounted for.</i> Cells are collective parts of the organ, not gross parts.
4.	Heart cells are parts of the heart, and the heart is part of the circulatory system, but yet the cells are not parts of the circulatory system.	<i>Accounted for.</i> Cells are collective parts of the Heart, not gross parts.

5.	Person P is part (member) of the football club FC, and FC is part (member) of the National Association of Football Clubs, NAFC, but yet P is not a part (member) of NAFC.	<i>Accounted for.</i> The person is a grain (member) of the football club, not a part of it and, similarly, the football club is a grain (member) of the association.
6.	Simpson's finger is part of Simpson, and Simpson is part of the Philosophy Department, but yet Simpson's finger is not part of the Philosophy Department.	<i>Accounted for.</i> Simpson is a member (grain) of the philosophy department (or possibly in some other relation to it), but not "part" of it in the sense used here.
7.	Hydrogen is part of water, and water is part of our cooling system, but yet hydrogen is not part of our cooling system.	<i>Accounted for</i> and a false example. Hydrogen is not part of water. Hydrogen atoms are part of water molecules, collectives of which constitute water used in the cooling system
8.	Cellulose is part of trees, and trees are parts of forests, but yet cellulose is not part of forests.	<i>Accounted for.</i> Trees are grains for forests.
9.	A handle is part of a spoon, and a spoon is part of eating soup, but yet a handle is not part of eating soup.	<i>Not accounted for;</i> A different issue. Continuants and occurents cannot be parts of each other for reasons not discussed in this paper.
10.	This shard was part of a plate, and the plate was part of a dinner service, but yet the shard was not part of the dinner service.	<i>Probably accounted for.</i> Leaving aside issues of time and the status of the shard prior to the breakable, the plate might be considered part of a collective.
11.	This tree is part of the Black forest, and the Black forest is part of Germany, but yet this tree is not part of Germany.	<i>Accounted for.</i> Trees are grains of forests. (Also the notion of geographical parthood might be treated differently by some authors)
12.	These grains of sand are part of the beach, and the beach is part of the island, but yet these grains of sand are not part of the island	<i>Accounted for.</i> The grains of sand are grains of the beach.

**Table 5: Johansson's list of cases for non-transitivity of part-whole relations**

We would argue that cases 4)-8) and 11)-12) are clearly accounted for by the distinction between gross and collective parthood.

Of the remainder, for cases 1 and 2, Johansson puts forward the argument that there is a narrow, non-transitive subproperty of parthood, which we usually term "direct parthood", that is not transitive and that the problem arises out of a confusion of the direct subproperty and the parent transitive property. He draws support for this distinction from Simons [20] and Casati & Varzi [5]. This seems to us entirely correct. However, Johansson also includes case 3 in this category. We would argue that it was better accounted for by the distinction between gross and collective parthood.

Case 9) Johansson explains by noting that two notions of parthood being used are fundamentally different. Again we would agree, a point we would signify by the incompatibility of parthood for occurents and continuants, *i.e.* "eating" and "spoon".

Case 10 is dealt with cursorily but seems clearly to raise a host of questions, not least whether the shard *per se* existed prior to the shattering of the plate.

Thus of Johansson's twelve cases, at least eight can be accounted for by the distinction between determinate and collective parthood; two are accounted for by confusion of direct and transitive parthood. Of the remainder, one is accounted for by the distinction between parthood for occurents and continuants (which is outside the scope of this paper), and one (case 10) is decidedly peculiar on several fronts. Johansson's thesis is that intransitive parthood predicates are not binary predicates. Our argument is that for the cases where it applies, the distinction between gross and collective parthood – *i.e.* between parthood without levels of collectivity rather than across them – is simpler, easier to apply, and arguably more fundamental.

### 3.4.3 Persistent and non-persistent part-hood

It is a general pattern that things continue to be spoken of as 'parts' even after they have been separated from the whole. Thus we speak of "John's finger" even after it has been amputated. Even if it has failed to develop we may speak of it as being absent. By contrast, we do not speak of the secretions from an individual cell as remaining part of that cell, although we might speak of them as being from one or more organs or even parts of organs. Hence we might legitimately seek to distinguish, for example, testosterone produced by the adrenal gland from testosterone produced by the testes, or oestrogen from the ovary from oestrogen from adipose tissue. However, we would be unlikely to distinguish testosterone originating from individual cells. Likewise, although we might talk of "cells from John's liver" following a biopsy, we would be unlikely to consider these as parts of John or his liver, present or missing, in the same sense as we would his amputated finger.

As in the above cases, we would argue that “persistent parthood” is something that pertains to things arising from gross parts but not from granular parts. This point, we accept, remains somewhat speculative and requires further investigation. (Note, we find “persistent parthood” as used here closer to common clinical usage than “permanent parthood” as advocated in Smith *et al.*) [24].

## 4. Discussion

### 4.1 Biomedical cases

#### 4.1.1 Tissues and substances

A major motivation for the current work is to deal with specific problems in the adequate representation of the biological notions of tissue and substance. In this formulation both are “mixtures” some of whose “ingredients” are “collectives”.<sup>9</sup> The schemas offered here provide both for properties that are intrinsic to the collective – *e.g.* arrangements and patterns – and for properties of the relation of the collective to the rest of the mixture, *e.g.* the proportion, distribution, etc. The claim is not that tissues are collectives, but that they are best viewed as amounts of matter some of whose ingredients are collectives. Much work remains to be done to describe patterns within tissues, but the schemas given provide a starting point. Whether the limitation of OWL and related languages to binary predicates will prove a major barrier remains to be seen. At this stage it is postulated that they are “good enough” for classification, but that further representation and inference mechanisms are likely to be required for applications such as detailed representation of developmental morphology.

#### 4.1.2 Why do current bio ontologies not make the distinction between granular and determinate parts?

An obvious question is: “If the distinction between determinate and granular parthood is so important, why is it not already standard?” The simplest answer is that few of the large bio-ontologies built to date have been required or used to support inferences that require this distinction.

In the Foundational Model of Anatomy [18, 25], the distinction is prefigured by the notion of “constituent parts”. However, the FMA is based exclusively on structure rather than function, so that the issue of emergent effects does not arise. Even when dealing with structure, the FMA does not represent attributes that apply to collectives such as the alignment of cells in the mucosa of the intestine (although the example is due to Cornelius Rosse<sup>10</sup>). Likewise, the FMA does not support detailed cardinality with respect to parts, so the distinction between fixed numbers of parts – *e.g.* fingers – and indeterminate numbers of parts – *e.g.* cells – does not arise. However, these limitations do present difficulties. The issue of the status of tissues and their structure is a significant problem and has, for example, plagued discussions in the SAEL consortium<sup>11</sup> in its efforts to reconcile various anatomic representations in mouse and man. The notions in this paper provide a framework for representing a number of the important notions raised in those discussions and a route towards reconciliation of some of the controversies.

In principle, the *OpenGALEN* ontology supports the distinction between collectives (termed “multiples”) and determinate parts (termed “components”). However, in practice it has often been elided. The prime use for *OpenGALEN* has been for defining surgical procedures and the drug actions and usages. In the first case attention is confined to determinate parts; in the second almost exclusively to granular parts (*e.g.* receptors). In very few cases is their room for confusion; hence the lack of distinction has not proved troublesome. Were the *OpenGALEN* model to be extended to include stronger modelling of physiology and function, then it is almost certain that the distinctions presented in this paper would become critical.

In SNOMED-CT, the primary use for anatomy is for the site, or locus, of diseases and the target of surgical and other interventions. Both uses are predominantly on the level of gross anatomy where collective effects are uncommon. Although notions such as “hair loss” must be defined as being literally “loss of at least one hair”, in practice no inferences turn on the detailed representation.

Does this neglect of the distinction between determinate and granular parts mean that the distinction is purely “academic”? We believe not. It merely reflects the current state of the art whereby representations are typically restricted to a single level of “collectivity”, or if you prefer, “granularity”.

As the demand for stronger functional representation across “levels of granularity” grows, including through the interoperation of extant ‘single level’ ontologies, so too will the need for a precise language to describe individual and collective effects and to distinguish them clearly from effects of physical size.

---

<sup>9</sup> The label “ingredient” is perhaps not ideal here. No better has yet been suggested, but the authors are open to suggestions.

<sup>10</sup> Private communication, 2004.

<sup>11</sup> <http://www.sofg.org/sael/>

## 4.2 Collectives and Normalisation of Ontologies

To support modularisation and maintenance, a major goal of the *OpenGALEN* ontologies is to maintain a “normalised” structure in their implementation in which all primitives form disjoint trees and all multiple classification is the result of inference rather than assertion [13]. The schema put forward here all lend themselves to normalisation in this sense. At least in its cognitivist/multiplicative versions, the different aspects of each entity are clearly factored so that they can be described independently.

## 4.3 Unresolved Issues

### 4.3.1 Determinate collectives

“Collectives” as discussed in this paper have an indeterminate number of grains. There are, however, collective effects of determinate collections of entities – the collective grip of the fingers, acuity of the eyes, the total capacity of the plates in a dinner service, etc. Note that in each of these cases, the collective effect is not determined by the precise number in the collective even though there is a ‘normative’ number. For example, a grip has strength whether one or more fingers is missing (or indeed a supernumerary finger were present), a person’s visual acuity is typically recorded whether a person has one or two functioning eyes, as being the best visual acuity with all the available eyes.

Formally, in the ontology, this presents no obvious problems. There is nothing to prevent the same entity being both a granular and a determinate part of the same or different wholes. Nonetheless, the issue of “determinate collectives” requires further exploration.

### 4.3.2 Empty collectives

This formulation allows for empty collectives, making it easy to represent proportions of zero. This poses no problem for the cognitivist view, but may pose some issue for a realist view. It is an open question how, in a realist perspective, such empty collectives are to be dealt with while avoiding a proliferation of “trivial” collectives that are part of everything.

### 4.3.3 Operations on Collectives

The most common requirement for operations on collectives is for variants of union and flattening. The collective of members of several collectives – *e.g.* the cells in the skin of the thumb and forefinger – can be easily expressed. Likewise, where collectives are nested, the flattened version can be easily captured – *e.g.* the collective of all cells in the collective of pancreatic eyelets. Although logically possible, the authors know of no application requiring such intersections of collectives.

### 4.3.4 Realist interpretation

The version presented here and in the accompanying OWL models is based on DOLCE which comes from the cognitivist/multiplicative tradition. Although the authors see no obstacles in principle to dealing with the realist formulation, there are significant technical issues and a proof of concept is not available at this time.

### 4.3.5 Are collectives of physical entities physical? material?

Whether collectives of physical entities should or should not count as physical has been deliberately left open in this paper. Likewise, whether empty collectives should be material. Because the schema for collectives uses “only” (*allValuesFrom*) rather than “some” (*someValuesFrom*) it is perfectly reasonable to assert axioms of the form, for example, that “all collectives of only physical entities are physical”, “all non-empty collectives of only physical entities are material.” These axioms seem both natural and helpful in biological applications. Similarly, it seems natural to treat non-empty collectives as material, and empty collectives as non-material, analogous to holes.

## 4.4 Conclusion: Providing a language for making distinctions clear

The word “granularity” has been used in so many different ways in so many different contexts that to try to enforce a single meaning on the term seems unlikely to succeed. We have therefore used the words “collectivity” and “size range” to distinguish two notions that are often lumped together under the general heading of “granularity” and argued that the two notions are effectively independent: that whereas collectives are necessarily larger than their grains, boundaries between levels of collectivity occur at all size ranges.

In general “physical objects” at one level form collectives that are “amounts of matter” at the next. Using more familiar but less precise language adapted from linguistics, countable entities at one level collectively form mass entities at the next. However, at least in biomedical applications, it is frequently necessary to refer alternatively to individual grains and the collectives that they form – *e.g.* both to “the sperm in the seminal fluid” and “the individual sperm that fertilises the egg”. There are also collectives that are themselves discrete and countable – *e.g.* a flock of geese, or a galaxy.

We have labelled the relation between grain and collective is `_grain_of` rather than the more familiar `is_member_of` to avoid confusion with mathematical sets defined extensionally.

In an area where the language is fraught, we invite alternative suggestions for the labelling of any of the notions in this paper. However labelled, we suggest that the central notion of collectives and grains is ubiquitous and accounts for important phenomena both in biomedical and broader ontologies and accounts for the criteria set out in the introduction in Section 1.3.

Although our expertise is in biomedicine, there is little specifically biological in these notions, and seems likely that these notions have much wider applicability.

## Acknowledgements

This work supported in part by the Semantic Mining Network of Excellence ((NoE 507505) sponsored by the European Commission, by the CLEF project (G011852) sponsored by the UK Medical Research Council, and by the CO-ODE/HyOntUse (GR/S44686/1) projects sponsored jointly by the UK Joint Infrastructure Services Committee (JISC) and UK Engineering and Physical Sciences Research Council (EPSRC).

## References

1. Artale, A., Franconi, E. and Guarino, N., Open problems for part-whole relations. in *International Workshop on Description Logics*, (Boston, MA, 1996), <http://www.dl.kr.org/dl96/>.
2. Artale, A., Franconi, E. and Pazzi, L. Part-whole relations in object-centered systems: An overview. *Data and Knowledge Engineering*, 20. 347-383.
3. Bada, M., Turi, D., McEntire, R. and Stevens, R. Using reasoning to guide annotation with gene ontology terms in GOAT. *ACM SIGMOD Record*, 33 (2). 27-32.
4. Bittner, T. and Smith, B. A theory of granular partitions. in Duckham, M., Goodchild, M. and Worboys, M. eds. *Foundations of Geographic Information Science*, Taylor & Francis Books, London, 2003, 117-151.
5. Casati, R. and Varzi, A.C. *Parts and Places*. Clarendon Press, Oxford, 1999.
6. Guarino, N. and Welty, C. An overview of OntoClean. in Staab, S. and Studer, R. eds. *sHandbook on Ontologies*, Springer Verlag, 2004, 151-159.
7. 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.
8. 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.
9. Masolo, C., Borgo, S., Gangemi, A., Guarino, N. and Oltraman, A. WonderWeb Deliverable 18, WonderWeb consortium, 2003, <http://www.loa-cnr.it/Papers/D18.pdf>
10. Mejino, J.L.V. and Rosse, C. Conceptualization of anatomical spatial entities in the Digital Anatomist Foundation Model. *Journal of the American Medical Informatics Association* (1999 Annual Symposium Special Issue). 112-116.
11. Odell, J.J. Six different kinds of composition. *Journal of Object Oriented Programming*, 5 (8). 10-15.
12. Padgham, L. and Lambrix, P., A Framework for Part-of Hierarchies in Terminological Logics. in *KR-94*, (1994), 485-496.
13. 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.
14. Rector, A., Gangemi, A., Galeazzi, E., Glowinski, A. and Rossi-Mori, A., The GALEN CORE Model Schemata for Anatomy: Towards a re-usable application-independent model of medical concepts. in *Twelfth International Congress of the European Federation for Medical Informatics, MIE-94*, (Lisbon, Portugal, 1994), 229-233.
15. 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.
16. Rogers, J. and Rector, A., The GALEN ontology. in *Medical Informatics Europe (MIE 96)*, (Copenhagen, 1996), IOS Press, 174-178.
17. Rogers, J. and Rector, A. GALEN's model of parts and wholes: Experience and comparisons. *Journal of the American Medical Informatics Association* ((Fall symposium special issue)). 819-823.

18. 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.
19. Schulz, S. and Hahn, U., Parts, locations, and holes - Formal reasoning about anatomical structures. in *Artificial Intelligence in Medicine Europe (AIME-2001)*, (Cascais, Portugal, 2001), Springer, 293-303.
20. Simons, P. *Parts. A study in Ontology*. Clarendon, Oxford, 1987.
21. Smith, B., The basic tools of formal ontology. in *Formal Ontology in Information Systems (FOIS)*, (Amsterdam, 1998), IOS Press (Frontiers in Artificial Intelligence and Applications), 19-28.
22. Smith, B. Granular partitions and vagueness. in Welty, C. and Smith, B. eds. *Formal Ontology and Information Systems*, Academic Press, New York, 2001, 225-237.
23. Smith, B., The logic of biological classification and the foundations of biomedical ontology. in *10th International Conference in Logic Methodology and Philosophy of Science*, (Oviedo Spain, 2004), Elsevier-North-Holland.
24. Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. and Rosse, C. Relations in Biomedical Ontologies. *Genome Biology*. (in press).
25. Smith, B. and Rosse, C., The role of foundational relations in the alignment of biomedical ontologies. in *Medinfo-2004*, (San Francisco, CA, 2004), IMIA, IOS Press, 444-448.
26. The Gene Ontology Consortium Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25. 25-29.
27. Welty, C. and Guarino, N. Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering*, 39 (1). 51-74.
28. Winston, M., Chaffin, R. and Hermann, D. A taxonomy of part-whole relations. *Cognitive Science*, 11. 417-444.
29. Wroe, C., Stevens, R., Goble, C.A. and Ashburner, M., An Evolutionary Methodology To Migrate The Gene Ontology To A Description Logic Environment Using DAML+OIL. in *Proceedings of the 8th Pacific Symposium on Biocomputing (PSB)*, (Hawaii, 2003), 624-635.