Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Biomedical Informatics 39 (2006) 333–349

Journal of Biomedical Informatics

www.elsevier.com/locate/yjbin

Granularity, scale and collectivity: When size does and does not matter

Alan Rector [a,*], Jeremy Rogers [a], Thomas Bittner [b]

[a] *Department of Computer Science, University of Manchester, Manchester M13 9PL, UK*
[b] *Institute of Formal Ontology and Medical Information Science, Saarland University, Germany*

Received 14 June 2005
Available online 28 November 2005

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 biomedical notion of parthood is, or is not, best represented by a transitive relation. Although examples are taken from biomedicine, we believe these notions to have wider application.

Keywords: Ontologies; Knowledge representation; Terminology; Part–whole relations

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.[1] The basic notion that we put forward is that entities considered individually at one level are considered as collectives with emergent properties 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."[2] The notion of "collective" used here is similar to that of "groups" used by Artale [1,2] and by Winston and Odell [3,4]. Winston and Odell also put forward an analogous line of argument to what are here called granular parts in discussing why the "feet of geese" are not parts of a "flock of geese." However, neither they nor Padgham and Lambrix [5] investigate this notion extensively. No analogous notion is discussed by authors such as Gerstl and Pribennow who discuss parts and wholes from a more linguistic perspective [6], nor do notions analogous to "collectives" and

* Corresponding author. Fax: +44 161 275 6204.
E-mail addresses: rector@cs.man.ac.uk (A. Rector), jrogers@cs.man.ac.uk (J. Rogers), thomas.bittner@ifomis.uni-saarland.de (T. Bittner).

[1] 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."

[2] Alternatively we might refer to collectives as "emergent wholes," but we have avoided this usage as collectives are usually themselves parts of greater wholes leading to awkward expressions such as "the emergent whole that is part of the whole."

1532-0464/$ - see front matter
doi:10.1016/j.jbi.2005.08.010

"granular parts" figure in the foundational relations discussed by Smith et al. [7]. In biomedical ontologies, the notion of "granular parts" is hinted at by the distinction between "constituent parts" and other forms of part–whole relation in the Foundational Model of Anatomy [8], but it is not extensively developed or explored. Overall, 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 and do not depend on differentiation amongst the properties of the grains. By "emergent" we mean that (a) these properties and effects cannot be predicted from the properties of the individual grains and therefore must be attributed to the collective as a whole, and that (b) all grains play the same role with respect to these properties and effects in the collective. Some properties only make sense of a collective—e.g., the pattern of a tiling or the arrangement of cells in a tissue. It makes no sense to speak of the pattern of a single tile or the alignment of a single cell. In other cases the emergent properties are distinct from that of the grains even if related, e.g., the mood of a crowd is distinct from the mood of its constituent individuals, a beach has area and galaxies have mass independent of the size of the grains of sand or the mass of the stars in the galaxy; tissues have strength, grow, etc., in ways distinct from the strength, growth, etc., of the individual cells that comprise them. The fundamental point is that properties of the whole and the information about it pertain to and are determined by the collective rather than its grains. Here we take as our prototype a classic hourglass. In some idealised world it might be possible to determine how long it took the sand to pass through an hourglass by examining the glass and the individual grains of sand and their initial configuration. In practice, no one would attempt such a feat. The time required for the sand to flow through the hourglass is a collective property of the sand in relation to the specific hourglass that contains it and would be measured as such. Even were someone, say a physicist specialised in fluid mechanics, to attempt such a feat, the 'gold standard' would remain the observed time—i.e., the emergent property of the collective.

Although the phenomenon of emergence is widely applicable, our fundamental motivations are biological. We seek:

1. 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 does not necessarily diminish the body whereas the loss of a finger does;
2. To use this to motivate an important criterion for when parthood as used in biomedicine should, or should not, be represented by a transitive relation;
3. To represent loosely repetitive patterns in tissues—that the "cells in the mucosa are aligned"—and more generally patterns and other emergent properties of collectives;
4. 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 entities.[3] Galaxies are more than mere collectives of stars; tissues are more than collectives of cells; even a beach is more than a collective of sand. If we 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, e.g., 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 *Open*GALEN Common Reference Model [9,10]. As an illustration we present this paper and an implementation in the framework of OWL-DL.[4] However, the issues are general and independent of any particular implementation.

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.

Furthermore we distinguish two types of parthood as subrelations of the basic mereological part–whole relation related to collectivity.

[3] Hence our reluctance to use the phrase "emergent whole" (See Footnote 2).

[4] An OWL-DL ontology illustrating the principles can be found at http://www.cs.man.ac.uk/~rector/ontologies/collectivity.