Chapter 1: Philosophy and Biomedical Information Systems

Barry Smith and Bert Klagges

1. *The New Applied Ontology*

Recent years have seen the development of new applications of the ancient science of philosophy, and the new sub-branch of *applied philosophy.* A new level of interaction between philosophy and non-philosophical disciplines is being realized. Serious philosophical engagement, for example, with biomedical and bioethical issues increasingly requires a genuine familiarity with the relevant biological and medical facts. The simple presentation of philosophical theories and arguments is not a sufficient basis for future work in these areas. Philosophers working on questions of medical ethics and bioethics must not only familiarize themselves with the domains of biology and medicine, they must also find a way to integrate the content of these domains in their philosophical theories. It is in this context that we should understand the developments in *applied ontology* set forth in this volume.

Applied ontology is a branch of applied philosophy using philosophical ideas and methods from ontology in order to contribute to a more adequate presentation of the results of scientific research. The need for such a discipline has much to do with the increasing importance of computer and information science technology to research in the natural sciences (Smith, 2003, 155-166). As early as the 1970s, in the context of attempts at data integration, it was recognized that many different information systems had developed over the course of time. Each system developed its own principles of terminology and categorization which were often in conflict with those of other systems. It was for this reason that a discipline known as ontological engineering has arisen in the field of information science whose aim, ideally conceived, is to create a common basis of communication – a sort of Esperanto for databases – the goal of which would be to improve the compatibility and reusability of electronically stored information.

Various institutions have sprung up, including the Metaphysics Lab at Stanford University, the Ontology Research Group in Buffalo, New York, and the Laboratories for Applied Ontology in Trento, Italy. Research at these institutions is focused on the use of ontological ideas and methods in

the interaction between philosophy and various fields of information sciences. The results of this research have been incorporated into software applications produced by technology companies such as Ingenuity Systems (Mountain View, California), Cycorp, Inc. (Austin, Texas), and Ontology Works (Baltimore, Maryland). Rapid developments in information-based research technology have called forth an ontological perspective, especially in the field of biomedicine. This is illustrated in the work of research groups and institutions such as Medical Ontology Research at the US National Library of Medicine, the Berkeley Bioinformatics and Ontology Project at the Lawrence Livermore National Laboratory, the Cooperative Ontologies Programme of the University of Manchester, the Institute for Formal Ontology and Medical Information Science (IFOMIS) in Saarbrücken, Germany, and the Gene Ontology Consortium.

2. *The Historical Background of Applied Ontology*

The roots of applied ontology stretch back to Aristotle (384-322 BCE), and from the basic idea that it is possible to obtain philosophical understanding of aspects of reality which are at the same time objects of scientific research.

But how can this old idea be endowed with new life today? In order to answer this question, we must cast a quick glance back at the history of Western philosophy. An ontology can be seen, roughly, as a taxonomy of entities – objects, attributes, processes, and relations – in a given domain, complete with formal rules that govern the taxonomy (for a detailed exposition, see Chapter 2). An ontology divides a domain into *classes* or *kinds* (in the terminology of this volume, *universals*). Complex domains require multiple levels of hierarchically organized classes. Carl Linnaeus's taxonomies of organisms are examples of ontologies in this sense. Linnaeus also applied the Aristotelian methodology in medicine by creating hierarchical categories for the classification of diseases.

Aristotle himself believed that reality in its entirety could be represented with one single system of categories (see Chapter 8). Under the influence of René Descartes and Immanuel Kant, however, the focal point of philosophy shifted from (Aristotelian) *metaphysics* to *epistemology*. In a separate development, the Aristotelian-inspired view of categories, species, and genera as parts of a determined order came gradually to be undermined within biology by the Darwinian revolution. In the first half of the twentieth century, this two-pronged anti-ontological turn received

increasing impetus with the influence of the logical positivism of the Vienna Circle.

Toward the end of the twentieth century, however, there was another shift of ground, in philosophy as well as in biology. Philosophers such as Saul Kripke, Hilary Putnam, David Armstrong, Roderick Chisholm, David Lewis, and Ruth Millikan managed to bring ontological and metaphysical considerations back into the limelight of analytic philosophy under the title 'analytical metaphysics'. This advance has brought elements of a still recognizably Aristotelian theory of categories (as the theory of universals or natural kinds) to renewed prominence. In addition, the growing importance of the new bioethics is helping to cast a new, ontological light on the philosophy of biology, above all in Germany in the work of Nikolaus Knoepffler and Ralf Stoecker.

In biology itself, traditional ideas about categorization which had been viewed as obsolete are now looked upon with favor once again. The growing significance of taxonomy and terminology in the context of current information-based biological research has created a terrain in which these ideas have blossomed once more. In fact, biology can be said to be enjoying a new golden age of classification.

3. *Ontological Perspectivalism*

One aspect of the Aristotelian view of reality still embraced by some ontologists is now commonly considered unacceptable, namely, that the whole of reality can be encompassed within one single system of categories. Instead, it is assumed that a multiplicity of ontologies – of partial category systems – is needed in order to encompass the various aspects of reality represented in diverse areas of scientific research. Each partial category system will divide its domain into classes, types, groupings, or kinds, in a manner analogous to the way in which Linnaeus's taxonomies divided the domain of organisms into various upper-level categories (kingdom, phylum, class, species, and so forth), now codified in works such as the *International Code of Zoological Nomenclature* and the *International Code of Nomenclature of Bacteria*.

One and the same cross-section of reality can often be represented by various divisions which may overlap with one another. For example, the Periodic Table of the Elements is a division of (almost) all of material reality into its chemical components. In addition, the table of astronomical categories, a taxonomy of solar systems, planets, moons, asteroids, and so

forth, is a division of (the known) material reality – but from another perspective and at another level of granularity.

The thesis that there are multiple, equally valid and overlapping divisions of reality may be called *ontological perspectivalism* (see Chapter 6). In contrast to various perspectival positions in the history of Western philosophy – for example, those of Nietzsche or Foucault – this ontological variant of perspectivalism is completely compatible with the scientific view of the world. Ontological perspectivalism accepts that there are alternative views of reality, and that this same reality can be represented in different ways. The same section of the world can be observed through a telescope, with the naked eye, or through a microscope. Analogously, the objects of scientific research may be equally well-viewed or represented by means of a taxonomy, theory, or language.

However, the ontological perspectivalist is confronted with a difficult problem. How can these various perspectives be made compatible with one another? How can scientific disciplines communicate, and work together, if each treats of a different subdivision or granularity? Is there a discipline which can provide some platform for integration? In the following we will try to show that, in tackling this problem, there is no alternative to an ontology constructed from philosophically grounded, rigorous formal principles. Our task is practical in nature, and is subject to the same practical constraints faced in all scientific activity. Thus, even an ontology based on philosophical principles always will be a partial and imperfect edifice, which will be subject to correction and enhancement, so as to meet new scientific needs.

4. *The Modular Structure of the Biological Domain*

The perspectives relevant to our purposes in the domain of biomedical ontology are those which help us to formulate scientific explanations. These are often perspectives of a fine granularity, by means of which we gain insight into, for example, the number and order of genes on a chromosome, or the reactions within a chemical pathway. But if the scientific view of these structures is to have a significance for the goals of medicine, it must be seen through different, coarse-grained perspectives, including the perspective of everyday experience, which embraces entities such as diseases and their symptoms, human feelings and behavior, and the environments in which humans live and act.

As Gottfried Leibniz asserted in the seventeenth century, when perceived more closely than the naked eye allows, the entities of the natural world are revealed to be aggregates of smaller parts. For example, an embryo is composed of a hierarchical nesting of organs, cells, molecules, atoms, and subatomic parts. The ecological psychologist Roger Barker expresses it this way:

> A unit in the middle range of a nesting structure is simultaneously both circumjacent and interjacent, both whole and part, both entity and environment. An organ – the liver, for example – is whole in relation to its own component pattern of cells, and is a part in relation to the circumjacent organism that it, with other organs, composes; it forms the environment of its cells, and is, itself, environed by the organism. (Barker, 1968, 154; compare Gibson, 1979)

Biological reality appears, in this way, as a complex hierarchy of nested levels. Molecules are parts of collections which we call cells, while cells are embedded, for example, in leaves, leaves in trees, trees in forests, and so forth. In the same way that our perceptions and behavior are more or less perfectly directed toward the level of our everyday experience, so too, the diverse biological sciences are directed toward various other levels within these complex hierarchies. There is, for example, not only clinical physiology, but also cell and molecular physiology; beside neuroanatomy there is also neurochemistry; and beside macroscopic anatomy with its sub-disciplines such as clinical, surgical, and radiological anatomy, there is also microscopic anatomy with sub-disciplines such as histology and cytology.

Ontological perspectivalism, then, should provide a synoptic framework in which the domains of these various disciplines can be linked, not only with each other, but also with an ontology of the granular level of the everyday objects and processes of our daily environment.

5. *Communication among Perspectives*

The central question is this: how do the coarse-grained parts and structures of reality, to which our direct perception and actions are targeted, relate to those finer-grained parts, dimensions, and structures of reality to which our scientific and technological capabilities provide access? This question recalls the project of the philosopher, Wilfrid Sellars, who sought what he called a stereoscopic view, the intent of which is to gather the content of our everyday thought and speech with the authoritative theories of the natural sciences into a single synoptic account of persons and the world

(Sellars, 1963). This stereoscopic view was intended to do justice, not only to the modern scientific image, but also to the manifest image of normal human reason, and to enable communication between them.

Which is the real sun? Is it that of the farmers or that of the astronomers? According to ontological perspectivalism, we need not decide in favor of the one or the other since both everyday and scientific knowledge stem from divisions which we can accept simultaneously, provided we are careful to observe their respective functions within thought and theory. The communicative framework which will enable us to navigate between these perspectives should provide a theoretical basis for treating one of the most important problems in current biomedicine. How do we integrate the knowledge that we have of objects and processes at the genetic (molecular) level of granularity with our knowledge of diseases and of individual human behavior, through to investigations of entire populations and societies?

Clearly, we cannot fully answer this question here. However, we will provide evidence that such a framework for integration can be developed as a result of the fact that biology and bioinformatics have implicitly come to accept certain theoretical and methodological presuppositions of philosophical ontology, presuppositions that pivot on an Aristotelian approach to hierarchical taxonomy.

Philosophical ideas about categories and taxonomies (and, as we will see, about many other traditional philosophical notions) have won a new relevance, especially for biology and bioinformatics. It seems that every branch of biology and medicine still uses taxonomic hierarchies as one foundation of its research. These include not only taxonomies of species and kinds of organisms and organs, but also of diseases, genomics and proteomics, cells and their components, biochemical reactions, and reaction pathways. These taxonomies are providing an indispensable instrument for new sorts of biological research in the form of massive databases such as Flybase, EMBL, Unigene, Swiss-Prot, SCOP, or the Protein Data Bank (PDB).[1] These allow new means of processing of data, resulting in extraction of information which can lead to new scientific results. Fruitful application of these new techniques requires, however, a solution to the problem of communication between these diverse category systems.

---

[1] See, for example, http://www.cs.man.ac.uk/~stevensr/ontology.html.

We believe that the new methods of applied ontology described in this volume bring us closer to a solution to this problem, and that it is possible to establish productive interdisciplinary work between biologists and information scientists wherein philosophers would act, in effect, as mediators.

6. *Ontology and Biomedicine*

There are many prominent examples of ways in which information technology can support biomedical research, including the coding of the human genome, studies of genetic expression, and better understanding of protein structures. In fact, all of these result from attempts to come to grips with the role of hereditary and environmental factors in health and the course of human diseases, and to search for material for new pharmaceuticals.

Current bioinformatics is extremely well-equipped to support calculation-intensive areas of biomedical research, focused on the level of the genome sequence, which can search for *quantitative* correlations, for example, through statistics-based methods for pattern recognition. However, an appropriate basis for *qualitative* research is less well-developed. In order to exploit the information we gain from quantitative correlations, we need to be able to process this information in such a way that we can identify those correlations which are of biological (and perhaps, clinical) significance. For this, however, we need a qualitative theory of types and relations of biological phenomena – an ontology – which also must include very general terms such as 'object', 'species', 'part', 'whole', 'function', 'process', and the like. Biologists have only a rather vague understanding of the meaning of these terms; but this suffices for their needs. Miscommunication between them is avoided simply in virtue of the fact that everyone knows which objects and processes in the laboratory are denoted by a given expression.

Information-technological processing requires explicit rigorous definitions. Such definitions can only be provided by an all-encompassing formal theory of the corresponding categories and relations. As noted already, information science has taken over the term 'ontology' to refer to such an all-encompassing theory. As is illustrated by the successes of the Gene Ontology (GO), developing such a resource can permit the mass of terminology and category systems thrown together in rather *ad hoc* ways over time to be unified within more overarching systems.

Already, the 1990s saw extensive efforts at modifying vocabularies in order to unite them within a common framework. Biomedical informatics offered framework approaches such as MeSH and SNOMED, as well as the creation of an overarching integration platform called the Unified Medical Language System (UMLS) (see National Library of Medicine). Little by little, the respective domains were indexed into robust and commonly accepted *controlled vocabularies*, and were annotated by experts to ensure the long-term compatibility and reusability of the electronically stored information. These controlled vocabularies contributed a great deal to the dawning of a new phase of terminological precision and orderliness in biomedical research, so that the integration of biological information that was hoped for seems achievable.

These efforts, however, were limited to the terminologies and the computer processes that worked with them. Much emphasis was placed upon the merely *syntactic* exactness of terms, that is, upon the grammatical rules applied to them as they are collected and ordered within structured systems. But too little attention was paid to the *semantic* clarity of these terms, that is, to their reference *in reality*. It was not that terms had no definitions – though such definitions, indeed, were often lacking. The problem was rather that these definitions had their origins in the medical dictionaries of an earlier time; they were written for people, not for computers. Because of this, they have an informal character, and are often circular and inconsistent. The vast majority of terminology systems today are still based on imprecisely formulated notions and unclear rules of classification.

When such terminologies are applied by people in possession of the requisite experience and knowledge, they deliver acceptable results. At the same time, they pose difficulties for the prospects of electronic data processing – or are simply inappropriate for this purpose. For this reason, the vast potential of information technology lies unexploited. For rigorously structured definitions are necessary conditions for consistent (and intelligent) navigation between different bodies of information by means of automated reasoning systems. While appropriately qualified, interested, and motivated people could make do with imprecisely expressed informational content, electronic information processing systems absolutely require exact and well-structured definitions (Smith, Köhler, Kumar, 2004, 79-94).

Collaboration between information scientists and biologists is all too often influenced by a variant of the *Star Trek* Prime Directive, namely,

'Thou shalt not interfere with the internal affairs of other civilizations'. In the present context, these other civilizations are the various branches of biology, while 'not to interfere' means that most information scientists see themselves as being obliged to treat information prepared by biologists as something untouchable, and so develop applications which enable navigation through this information. Hence, information scientists and biologists often do not interact during the process of structuring their information, even though such interaction would improve the potential power of information resources tremendously. Matters are changing, now, with the development of OBI, the OBO Foundry Ontology for Biomedical Investigations (http://obi.sourceforge.net/), which is designed to support the consistent annotation of biomedical investigations, regardless of the particular field of study.

7. *The Role of Philosophy*

Up to now, not even biological or medical information scientists were able to achieve an ontologically well-founded means of integrating their data. Previous attempts, such as the Semantic Network of the UMLS (McCray, 2003, 80-84), brought ever more obvious problems stemming from the neglect of philosophical, logical, and especially definition-theoretical principles for the development of ontological theories to light (Smith, 2004, 73-84). Terms have been confused with concepts, while concepts have been confused with the things denoted by the words themselves and with the procedures by which we obtain knowledge about these things. Blood pressure has been identified, for example, with the *measuring* of blood pressure. Bodily systems, such as the circulatory system, have been classified as conceptual entities, but their parts (such as the heart) as physical entities. Further, basic philosophical distinctions have been ignored. For example, although the Gene Ontology has a taxonomy for functions and another for processes, initially there was no attempt to understand how these two categories relate or differ; both were equated in GO with 'activity'. Recent GO documentation has improved matters considerably in these respects, with concomitant improvements in the quality of the ontology itself.

Since computer programs only communicate what has been explicitly programmed into them, communication between computer programs is more prone to certain kinds of mistakes than communication between people. People can read between the lines (so to speak), for example, by

drawing on contextual information to fill in gaps of meaning, whereas computers cannot. For this reason, computer-supported systems in biology and medicine are in dire need of maximal clarity and precision, particularly with respect to those most basic terms and relations used in all systems; for example, '*is_a*', '*part_of*', or '*located_in*'. An ontological theory based on logical and philosophical principles can, we believe, provide much of what is needed to supply this missing clarity and precision, and early evidence from the development of the OBO Foundry initiative is encouraging in this respect. This sort of ontological theory can not only support more coherent interpretations of the results delivered by computers, it also will enable better communication between, and among, the scientists of various disciplines. This is achieved by counteracting the fact that scientists bring a variety of different background assumptions to the table and, for this reason, often experience difficulties in communicating successfully.

One instrument for improving communication is the OBO Foundry's Foundational Model of Anatomy (FMA) Ontology, developed through the Department of Biological Structure at the University of Washington in Seattle, which is a standard-setter among bioinformation systems. The FMA represents the structural composition of the human body from the macromolecular level to the macroscopic level, and provides a robust and consistent schema for the classification of anatomical unities based upon explicit definitions. This schema also provides the basis for the Digital Anatomist, a computer-supported visualization of the human body, and provides a pattern for future systems to enable the exact representation of pathology, physiological functions, and the genotype-phenotype relations.

The anatomical information provided by the FMA Ontology is explicitly based upon Aristotelian ideas about the correct structure of definitions (Michael, Mejino, Rosse, 2001, 463-467). Thus, the definition of a given class in the FMA – for example, the definition for 'heart' or 'organ' – specifies what the corresponding instances have in common. It does this by specifying (a) a *genus*, that is, a class which encompasses the class being defined, together with (b) the *differentiae* which characterize these instances within the wider class and distinguish them from its other members. This modular structure of definitions in the FMA Ontology facilitates the processing of information and checking for mistakes, as well as the consistent expansion of the system as a whole. This modular structure also guarantees that the classes of the ontology form a genuine categorial tree in the ancient Aristotelian sense, as well as in the sense of the Linnaean taxonomy. The Aristotelian doctrine, according to which

definition occurs via the nearest genus and specific difference, is applied in this way to current biological knowledge.

In earlier times the question of which types or classes are to be included within the domain of scientific anatomy was answered on the basis of visual inspection. Today, this question is the object of empirical research within genetics, along with a series of related questions concerning, for example, the evolutionary predecessors of anatomical structures extant in organisms. In course of time, a phenomenologically recognizable anatomical structure is accepted as an instance of a genuine class by the FMA Ontology only after sufficient evidence is garnered for the existence of a structural gene.

8. *The Variety of Life Forms*

The ever more rapid advance in biological research brings with it a new understanding of the variety of characteristics exhibited by the most basic phenomena of life. On the one hand, there is a multiplicity of *substantial forms of life*, such as mitochondria, cells, organs, organ systems, single- and many-celled organisms, kinds, families, societies, populations, as well as embryos and other forms of life at various phases of development. On the other hand, there are certain basic building blocks of *processes,* what we might call *forms of processual life*, such as circulation, defence against pathogens, prenatal development, childhood, adolescence, aging, eating, growth, perception, reproduction, walking, dying, acting, communicating, learning, teaching, and the various types of social behavior. Finally, there are certain types of processes, such as cell division or the transport of molecules between cells, in every phase of biological development.

Developing a consistent system of ontological categories founded upon robust principles which can make these various forms of life, as well as the relations which link them, intelligible requires addressing several issues which are often ignored in biomedical information systems, or addressed in an unsatisfactory manner, because they are philosophical in nature. These issues show the unexplored practical relevance of philosophical research at the frontier between information science and empirical biology.[2] These issues include:

---

[2] See also: Smith, Williams, Schulze-Kremer, 2003, 609-613; Smith, Rosse, 2004, 444-448.

(1) Issues pertaining to the different modes of existence through time of diverse forms of life. Substances (for example, cells and organisms) are fundamentally different from processes with respect to their mode of existence in time. Substances exist as a whole at every point of their existence; they maintain their identity over time, which is itself of central relevance to the definition of 'life'. By contrast, processes exist in their temporal parts; they unfold over the course of time and are never existent as a whole at one and the same instant (Johansson, 1989; Grenon, Smith, 2004, 69-103).

We can distinguish between entities which exist continually (continuants) and entities which occur over time (occurrents). It is not only substances which exist continually, but also their states, dispositions, functions, and qualities. All of these latter entities stand in certain relations on the one hand to their substantial bearers and on the other hand to certain processes. For example, functions are generally *realized* in processes. In the same way that an organism has a *life*, a *disposition* has the possibility of being *realized*, and a *state* (such as a disease) has its *course* or its *history* (which can be represented in a medical record).

(2) The notion of *function* in biology also requires analysis. It is not only genes which have functions that are important for the life of an organism; so do organs and organ systems, as well as cells and cellular parts such as mitochondria or chloroplasts. A function inheres in a body part or trait of an organism and is realized in a process of *functioning*; hence, for example, one function of the heart is to pump blood. But what does the word 'function' mean in this context? Natural scientists and philosophers of science from the twentieth century have deliberately avoided talk of functions – and of any sort of teleology – because teleological theories were seen to be in disagreement with the contemporary scientific understanding of causation. Yet, functions are crucial for the worldview (the ontology) of physicians and medical researchers, as a complete account of a body part or trait often requires reference to a function. Further, it is in virtue of the body's ability to transform malfunctioning into functioning that life persists.

The nature of functions has been given extensive treatment in recent philosophy of biology. Ruth Millikan, for example, has offered a theory of proper function as a disposition belonging to an entity of a certain type, which developed over the course of evolution and is responsible for (at least in part) the existence of more entities of its type (Millikan, 1988). However, an entity has a function only within the context of a biological

system and this requires, of course, an analysis of *system*. But existing philosophical theories lack the requisite precision and general application necessary for a complete account of functions and systems (Smith, Papakin, Munn, 2004, 39-63; Johansson, *et al.*, 2005, 153-166).

(3) The issue of the components and structure of organisms also needs to be addressed. In what relation does an organism stand to its body parts? This question is a reappearance of the ancient problem of *form* and *matter* in the guise of the problem of the relation between the organism as an organized whole, and its various material bearers (nucleotides, proteins, lipids, sugars, and so forth). Single-celled as well as multi-celled organisms exhibit a certain modular structure, so that various parts of the organism may be identified at different granular levels. There are a variety of possible partitions through which an organism and its parts can be viewed depending upon whether one's focus is centered on molecular or cellular structures, tissues, organ systems, or complete organisms. Because an organism is more than the sum of its parts, this plurality of trans-granular perspectives is central to our understanding of an organism and its parts. The explanation of how these entities relate to one another from one granular level to the next is often discussed in the literature on emergence, but is seldom imbued with the sort of clarity needed for the purposes of automated information representation.

The temporal dimension contains modularity and corresponding levels of granularity as well. So, if we focus successively on seconds, years, or millennia, we perceive the various partitions of processual forms of life, such as individual chemical reactions, biochemical reaction paths, and the life cycles of individual organisms, generations, or evolutionary epochs.

(4) We also need to address the issue of the nature of biological kinds (species, types, universals). Any self-respecting theory of such entities must allow room for the *evolution of kinds*. Most current approaches to such a theory appeal to mathematical set theory, with more or less rigor. A biological kind, however, is by no means the same as the set of its instances. For, while the identity of a set is dependent upon its elements or members and, hence, participates to some degree in the world of time and change, sets themselves exist outside of time. By contrast, biological kinds exist in time, and they continue to exist even when the entirety of their instances changes. Thus, biological kinds have certain attributes in common with individuals (Hull, 1976, 174-191; Ghiselin, 1997), and this is an aspect of their ontology which has been given too little attention in bioinformatics.

Existing bioinformation systems concentrate on terms which are organized into highly general taxonomical hierarchies and, thus, deal with biological reality only at the level of classes (kinds, universals). Individual organisms – which are instantiations of the classes represented in these hierarchies – are not taken into consideration. This lack of consideration has partially to do with the fact that the medical terminology, which constitutes the basis for current biomedical ontologies, so overwhelmingly derives from the medical dictionaries of the past. Authors of dictionaries, as well as those involved in knowledge representation, are mainly interested in what is general. However, an adequate ontology of the biological domain must take individuals (instances, particulars) as well as classes into account (see Chapters 7, 8, and 10). It must, for example, do justice to the fact that biological kinds are always such as to manifest, not only typical instances, but also a penumbra of borderline cases whose existence sustains biological evolution. As we will show in what follows, if we want to avoid certain difficulties encountered by previous knowledge representation systems, the role of instances in the structuring of the biological domain cannot be ignored.

(5) There is much need, also, for a better understanding of synchronic and diachronic identity. Synchronic identity has to do with the question of whether x is the same individual (protein, gene, kind, or organism) as y, while diachronic identity concerns the question of whether x is *today* the same individual (protein, gene, kind, or organism) as x was yesterday or a thousand years ago. An important point of orientation on this topic is the logical analysis of various notions of identity put forward by the *Gestalt*-psychologist Kurt Lewin (Lewin, 1922). Lewin distinguishes between physical, biological, and evolution-theoretic identity; that is, between the modes of temporal persistence of a complex of molecules, of an organism, or of a kind. Contemporary analytic philosophers, such as Eric Olson or Jack Wilson, have also managed to treat old questions (such as those of personal identity and individuation) with new ontological precision (Olson, 1999; Wilson, 1999).

(6) There is also a need for a theory of the role of environments in biological systems. Genes exist and are realized only in very specific molecular contexts or environments, and their concrete expression is dependent upon the nature of these contexts. Analogously, organisms live in niches or environments particular to them, and their respective environments are a large part of what determine their continued existence.

However, the philosophical literature since Aristotle has shed little light upon questions relating to the ontology of the environment, generally according much greater significance to substances and their accidents (qualities, properties) than to the environments surrounding these substances. But what *are* niches or environments, and how are the dependence relations between organisms and their environments to be understood ontologically? The relevance of these questions lies not only within the field of developmental biology, but also ecology and environmental ethics, and is now being addressed by the OBO Foundry's new Environment Ontology (http://environmentontology.org).

9. *The Gene Ontology*

The rest of this volume will provide examples of the methods we are advocating for bringing clarity to the use of terms by biologists and by bioinformation systems. We will conclude this chapter with a discussion of the Gene Ontology (see Gene Ontology Consortium, ND), an automated taxonomical representation of the domains of genetics and molecular biology. Developed by biologists, the Gene Ontology (GO) is one of the best known and most comprehensive systems for representing information in the biological domain. It is now crucial for the continuing success of endeavors such as the Human Genome Project, which require extensive collaboration between biochemistry and genetics. Because of the huge volumes of data involved, such collaboration must be heavily supported by automated data exchange, and for this the controlled vocabulary provided by the GO has proved to be of vital importance.

By using humanly understandable terms as keys to link together highly divergent datasets, the GO is making a groundbreaking contribution to the integration of biological information, and its methodology is gradually being extended, through the OBO Foundry, to areas such as cross-species anatomy and infectious disease ontology.

The GO was conceived in 1998, and the Open Biomedical Ontologies Consortium (see OBO, ND) created in 2003, as an umbrella organization dedicated to the standardization and further development of ontologies on the basis of the GO's methodology. The GO includes three controlled vocabularies – namely, *cellular component*, *biological process*, and *molecular function* – comprising, in all, more than 20,000 biological terms. The GO is not itself an integration of databases, but rather a vocabulary of terms to be used in describing genes and gene products. Many powerful

tools for searching within the GO vocabulary and manipulation of GO-annotated data, such as AmiGO, QuickGO, GOAT, and GoPubMed (see GOAT, 2003 and gopubmed.org, 2007), have been made available. These tools help in the retrieval of information concerning genes and gene products annotated with GO terms that is not only relevant for theoretical understanding of biological processes, but also for clinical medicine and pharmacology.

The underlying idea is that the GO's terms and definitions should depend upon reference to individual species as little as possible. Its focus lies, particularly, on those biological categories – such as *cell*, *replication*, or *death* – which reappear in organisms of all types and in all phases of evolution. It is not a trivial accomplishment on the GO's part to have created a vocabulary for representing such high-level categories of the biological realm, and its success sustains our thesis that certain elements of a philosophical methodology, like the one present in the work of Aristotle, can be of practical importance in the natural sciences.

Initially, the GO was poorly structured and some of its most basic terms were not clearly defined, resulting in errors in the ontology itself. (See: Smith, Köhler, Kumar, 79-94; Smith, Williams, Schulze-Kremer, 609-613). The hierarchical organization of GO's three vocabularies was similarly marked by problematic inconsistencies, principally because the *is_a* and *part_of* relations used to define the architecture of these ontologies were not clearly defined (see Chapter 11).

In early versions of the GO, for example, the assertions such as 'cell component *part_of* Gene Ontology' existed alongside properly ontological assertions such as 'nucleolus *part_of* nuclear lumen' and 'nuclear lumen *is_a* cellular component'. Unlike the second and third assertions, which rightly relate to part-whole relations on the side of biological reality, the first assertion captures an inclusion relation between a term and a list of terms in the GO itself. This misuse of '*part_of*' represents a classic confusion of *use* and *mention*. A term is *used* if its meaning contributes to the meaning of the including sentence, and it is merely *mentioned* if it is referred to, say in quotation marks, without taking into account its meaning (for more on this distinction and its implications, see Chapter 13).

10. *Conclusion*

The level of philosophical sophistication among the developers of biomedical ontologies is increasing, and the characteristic errors by which

such ontologies were marked is decreasing as a consequence. Major initiatives, such as the OBO Foundry, are a reflection of this development, and further aspects of this development are outlined in the chapters which follow.