3 Principles of Best Practice I: Domain Ontology Design

In our Introduction, we articulated the problem of managing scientific information in a way that allows combinability and comparability, and we discussed ontology as a proposed general solution to this problem. In chapters 1 and 2, an ontology was defined as a representational artifact whose representations are intended to designate universals, defined classes, and the relations among them. We also introduced some distinctions among different kinds of ontologies, and introduced the idea of a taxonomy as the central component of an ontology. In light of all of this, the problem of designing an ontology is the problem of designing a formalized representational artifact, including a taxonomic hierarchy as backbone, whose representations (terms) designate universals, defined classes, and the relations between them. In this chapter and the next we will discuss what this process looks like in practice, focusing on considerations and principles geared toward the design of domain reference ontologies useful in supporting scientific research. Issues to be considered in this chapter include: subject matter and scope of a domain ontology, as well as the first steps that one should take in designing the ontology itself.

General Principles of Ontology Design

We will first set forth principles specifying the general attitude or outlook to be kept in mind when designing an ontology. Our position is that a good ontology will be one that is designed in such a way as to respect these principles and that, indeed, respecting these principles will be part of what makes an ontology a good one.

1. Realism

We have already discussed our commitment to realism in chapter 1. In general, "realism" can be defined as a philosophical position according to which reality and its constituents exist independently of our (linguistic, conceptual, theoretical, cultural) representations and can be known, for example, through perceptual experience and through application of the scientific method. The goal of science, from this realist (and, we believe, commonsensical) perspective, is to discover truths about reality.

Realism in ontology is based further on the idea that with the aid of science we can come to know the general features of reality in the form of universals and the relations between them. This realist approach has a number of general consequences. First, it implies that ontologies are representations of reality, not of people's concepts or mental representations or uses of language. Certainly an ontology of, for example, cognitive psychology or linguistics might have concepts or mental representations or uses of language within its subject matter. But then the latter would be treated as parts of reality in a way exactly analogous to what is the case in, for example, an ontology of astrophysics or plant development.

Many parts of science pertain to entities, such as chemical elements or prokaryote cells or Paleoproterozoic rocks, which existed long before the first human beings. Other parts of science pertain to entities—for example, in the domains of law or economics—which exist as a result of human thought and activity. Ontological realism applies equally to all branches of science, taking the view that, for example, collateralized debt obligations are no less real than electrons and planets.

2. Perspectivalism

The goal of science is not merely to discover truths about reality. Its goal is to develop theories that are as accurate, as broad ranging, as predictive, as explanatory, as logically coherent, and as well tested as possible. Unfortunately these goals—and a number of other goals that are also found attractive, such as maximal consistency with common sense—seem not to be simultaneously realizable. To cope with this fact, we embrace a doctrine of *perspectivalism*.

Perspectivalism flows from the recognition that reality is too complex and variegated to be embraced in its totality within a single scientific theory. It amounts to the principle that two distinct scientific theories may both be equally accurate representations of one and the same reality.

This does not mean, of course, that all representations created by scientists are of equal value. A view according to which fish are mammals would clearly be of less value than a contrary view, because it would be less accurate to the facts of reality. But there are nonetheless many different representations that are equally good (true, veridical) representations of some given portion of reality precisely because they capture different features of this reality. The most straightforward examples of different but equally legitimate perspectives regarding the same domain of reality have to do with the phenomenon of granularity. Simply put, it is equally legitimate to examine living organisms from a perspective of molecular biology as from a perspective that takes into account anatomy and physiology at the level of organs and organ systems. It is equally legitimate to examine human behavior from the perspective of the physics of the human sensorimotor system as from the perspective of economic incentives. Each of the mentioned perspectives can yield contributions to our knowledge of reality that are accurate to the reality at hand.

The implications of perspectivalism for ontology are that the irreducibility of different perspectives should be respected also in the design of ontologies. Ontology developers should not seek to represent all portions and features of reality in a single ontology, but should seek, rather, a modular approach, in which each module is maintained as far as possible by experts in the corresponding scientific discipline.

3. Fallibilism

Fallibilism involves commitment to the idea that, even though our current scientific theories are the best source we have of statements that are candidates for expressing truths about reality, it may nevertheless be the case that some of these statements are false. Reality exists independently of our ways of understanding it scientifically, and experience tells us that even our best current theories may be subject to correction. Thus while the realist holds that our experiences, ideas, and scientific theories are about reality—that they form in totality a representation, a map or picture, of reality—this does not mean that all elements of this map are correct; some elements may misrefer, some may fail to refer at all.

Our map of reality is at any given stage always only partial: reality is never revealed to scientists in all its totality. And because our representation is continually expanding as we learn and discover more about what exists on the side of reality, what we believe today, based on what we have learned about those facets of reality to which we have so far gained access, is sometimes undermined by what we learn tomorrow about facets of reality hitherto unappreciated.

The process of correction of our map of reality is itself subject to multiple different sorts of setbacks and changes of direction, some (few) of which may be radical in nature (two outstanding examples being the Copernican and Darwinian revolutions in physics and biology). Through all of these changes, however, and even through the most radical of scientific revolutions, major referential elements of this map remain intact. Scientists erred in believing that the sun rotates around the earth; but upon correcting this error they continued to use terms like “sun” and “earth” to refer to the very same entities as before. Something similar applies to general terms like “atom,” "star," "organism," "cell," and "planet." While our beliefs about these entities have changed with time, these terms themselves have to a large degree preserved their reference through such changes. At the same time, however, the fallibilist accepts that in regard to general terms, too, our scientific knowledge is subject to being overturned through time through new empirical discoveries, as, for example, in the already mentioned "phlogiston" case.

Some specific implications of fallibilism for ontology design in support of scientific research include the following:

3a. That every ontology must have sophisticated strategies for keeping track of successive versions of the ontology.

A new version of an ontology becomes necessary when errors in current scientific theory about the domain are discovered and corrected and when new information relevant to the domain is discovered. Users of the ontology need to be able to keep track of such changes.

3b. That every ontology needs to have a tracking service for its users that will allow them easily to point out errors and gaps in the ontology and to have their submissions to this service addressed in a timely manner.

Like science itself, ontology design is an ongoing collective enterprise in which errors can be detected and avoided by the input and testing of multiple individuals.

4. Adequatism

There is a widespread tendency in philosophical circles to view the goal of philosophy in reductionist terms. On this view the job of the philosopher is to explain complex phenomena by reducing them to simpler and more fundamental components, here drawing on the astonishing successes of modern physics. *Adequatism* is the opposite tendency, which holds that the entities in any given domain should be taken seriously on their own terms and that room must be made in our set of theories of reality for all of the different sorts of entities that reality contains, at all levels of granularity.

For the adequatist all scientific disciplines are prima facie of equal worth in providing representations of what exists in reality. Just as an ontology of physics is about, for example, atoms and subatomic particles, and an ontology of chemistry is about chemical elements and compounds and the associated reactions, so an ontology of biology will include representations of the universals and defined classes at various levels of granularity from molecules and cells, to organs and organisms, and from there to populations and ecosystems. The goal of ontology as viewed by the proponent of adequatism is to do justice to the vast array of different kinds of entities that exist in the world, rather than ignoring these or those specific kinds of entities or attempting to explain them away.

It is the adequatist view of ontology that is defended in what follows. Suppose, for example, that one needs to create an ontology for a given domain as this domain is described in the textbooks of some given scientific discipline. The ontology should be designed to represent the types of entities described in the textbooks; but it should do this in such a way that it can be linked to other ontologies covering neighboring domains, including domains recognizing entities at different levels of granularity. The implication is that ontologies should not be developed in isolation from each other, but rather always in tandem with ontologies with which they must interoperate.

More generally, an adequate framework for ontology development should allow for entities at multiple levels of granularity (as, for example, in biology, where an adequate general framework must allow for—at least—molecules, cells, organs, organisms, and populations) and for a variety of different kinds of relations between the entities on these different levels.

Additional Principles of Ontology Design

Whereas the foregoing four principles represent general theoretical standpoints about ontology design, the following four are more concrete guidelines concerning the design process itself.

5. The Principle of Reuse

Ontologists should not reinvent the wheel. The first step in ontology development should always be one of examining existing ontology resources in and around the domain of focus in order to identify content that is already available that meets scientific and ontological standards. Ontologies should reuse as far as possible relevant ontological content that has already been created; and even where this content cannot be reused it should be regarded as forming a benchmark that can be used to gauge the adequacy of new content that is created.

Ontologies are designed to support communication between information resources relating to the multiple provinces of reality and to the multiple disciplines which seek to describe them. They can be compared, in this respect, to highway systems. It will very rarely be the case that the correct solution to an ontological problem involves the equivalent of ignoring all the roads that already exist, and creating an entirely new highway system from scratch.

At the same time, however, it must be emphasized that—precisely because ontologists have so often ignored design principles of the sort presented here, and because they have themselves often created new ontologies from scratch—much available ontology content is poor in quality, and so due diligence is required not merely in identifying potential ontologies for reuse, but also in evaluating the ontologies identified (and in some cases recommending that they be excluded from further use).

6. The Ontology Design Process Should Balance Utility and Realism

It is an implication of realism that some representational schemes are better than others because they are better representations of reality. Given that some of the roots of ontology building lie in the field of what is sometimes called knowledge engineering, where highly practical motives predominate, it is often argued that ontologies should be measured not by this global standard of adequacy to reality—a standard adapted from the domain of science as a whole—but rather by their utility to some specific local purpose. From our point of view, however, this focus on local utility is wrongly conceived if it is seen as involving a sacrifice on the side of adequacy to the reality which the ontology is being constructed to represent. For it is this reality—as described in the best current science—which provides the common benchmark that can ensure that ontologies are developed in a consistent fashion. Ontologies can indeed be developed in the absence of such a benchmark, but then when they are used to annotate data, the results will not be combinable—except perhaps through considerable manual effort—with data collected by others in neighboring domains. One lesson from over fifteen years of experience with the Gene Ontology is that the purpose for which an ontology originally is constructed may differ in significant ways from what turn out to be important secondary uses that could not have been anticipated when the ontology was first conceived.

7. The Ontology Design Process Is Open-Ended

The principles discussed so far provide a framework for understanding a crucial further point about ontology design: designing a domain ontology, at least in the scientific domains of primary concern to us here, is but the first step in an open-ended process of continuously maintaining, evaluating, updating, and correcting the ontology, and of adjusting the ontology to neighboring ontologies in order to take account of advances both in scientific knowledge and in our knowledge of ontology and its associated logical and computational technologies.

Realism implies that the central goal of a good ontology in support of scientific research is to represent reality adequately. But it implies also that for scientific domains we are at any given stage almost always in possession of only partial information about the reality at issue. Our strategy thus forces upon us the precept that ontologies should be designed in such a way as to be expandable and amendable through time, and the best practice principles that we will be discussing in what follows are designed to serve this end. Note that this precept is consistent with the fact that there will be practical constraints on the ontology builder resulting from the fact that resources for populating an ontology are limited by economic and other circumstances. For while those branches of the ontology that are associated with the most urgent needs will be developed in greatest detail, the population of such branches will be of greater utility if it is managed within a general framework that can allow coherent population of neighboring branches in the future.

8. The Principle of Low-Hanging Fruit

A final general principle to keep in mind is the following: when designing a domain ontology, begin by identifying those features of the subject matter that are the easiest and clearest to understand and define.

The ontologist should begin, in other words, by gathering the low-hanging fruit on the ontology tree, including what to a human being may seem like trivial assertions (for example, *cell membrane is_a membrane*) but which to the computers who will process the ontology are indispensable. In constructing a domain ontology we start by categorizing the simple universals and relations first. As a general rule, the ontology developer should start by identifying the general terms most commonly used in the beginning chapters of relevant introductory textbooks, and move on from there, step by step, to the representation of more complex entities within the domain. The principles of ontology design that have been surveyed up to this point are summarized in box 3.1.

Overview of the Domain Ontology Design Process

Ontology is a top-down approach to the problem of electronically managing scientific information. This means that the ontologist begins with theoretical considerations of a very general nature on the basis of the assumption that keeping track of more specific information (for example, about specific organs, genes, or diseases) requires getting the very general scientific framework underlying this information right, and doing so in a systematic and coherent fashion. It is only when this has been done that the detailed terminological content of a specific science such as cell biology or immunology can be encoded in such a way as to ensure widespread accessibility and usability. The method to be followed in constructing a domain ontology on the basis of this general starting point can be summarized in the steps outlined in table 3.1.

Box 3.1
General Principles of Ontology Design

1. *Realism*: the goal of an ontology is to describe reality.
2. *Perspectivalism*: there are multiple accurate descriptions of reality.
3. *Fallibilism*: ontologies, like scientific theories, are revisable in light of new discoveries.
4. *Adequatism*: the entities in a given domain should be taken seriously on their own terms, not viewed as reducible to other kinds of entities.
5. *The Principle of Reuse*: existing ontologies should be treated as benchmarks and reused whenever possible in building ontologies for new domains.
6. *The Ontology Design Process Should Balance Utility and Realism*: sacrificing realism to address considerations of short-term utility when building an ontology may be at the detriment of the ontology's longer-term utility.
7. *The Ontology Design Process Is Open-Ended*: scientific ontologies will always be subject to the need for update in light of advances in knowledge; ontology design, maintenance, and updating is an ongoing process.
8. *The Principle of Low-Hanging Fruit*: in ontology design, begin with the features of the relevant domain that are easiest to understand and define, then work outward to more complex and controversial features.

Table 3.1
An outline of the steps to be followed in designing a domain ontology

1. Demarcate the subject matter of the ontology.
2. Gather information: identify the general terms used in existing ontologies and in standard textbooks; analyze to remove redundancies.
3. Order these terms in a hierarchy of the more and less general ones.
4. Regiment the result in order to ensure: a. logical, philosophical, and scientific coherence, b. coherence and compatibility with neighboring ontologies, and c. human understandability, especially through the formulation of human-readable definitions.
5. Formalize the regimented representational artifact in a computer usable language in such a way that the result can be implemented in some computable framework.

Step 1 consists of determining and demarcating the subject matter of the ontology one needs to build. This will involve establishing the nature and scope of the data (for example, experimental or clinical) one needs to annotate, and identifying existing ontology content in relevant domains. The initial survey of the content of the pertinent science should yield provisional answers to the following questions:

- What are the domain universals and relations that need to be represented?
- What are the appropriate domain-specific terms that should be used in representing these universals and relations?
- What levels of granularity of entities are salient for the ontology?

Step 2 is the task of assembling a selection (fifty or so) of the most common highly general terms, some of them taken from relevant ontologies, some from standard textbooks.

Step 3 is the provisional ordering of these terms in a hierarchy of the more and less general and serves as the precursor to step 4. Step 4 consists in working on this hierarchy to ensure coherence, for example, by adding further terms to ensure a complete taxonomical hierarchy for the ontology; and identifying the terms referring to the highest-level universals in the domain in question, which will serve as the root node or nodes of the ontology being developed. It will involve also creating a set of human-understandable definitions for the selected terms, which will include gathering further information concerning the most important domain-specific universals which are subsumed by these highest-level universals, and identifying any relevant terms in neighboring ontologies which will be needed in the formulation of the definitions. Starting with the root nodes and working downward, we attempt to identify successive genera and differentiating characteristics that will need to be included in the definitions of the entities to be included in the ontology; and we adjust our preliminary classification scheme in light of any changes which our definitions dictate.

The process of regimentation is iterative, and will involve successive cycles of reviewing versions of the hierarchy of terms and definitions for logical, philosophical, and scientific adequacy, including consistency and human intelligibility, and also ensuring that the result leaves out no essential elements of the domain.

Once a thorough understanding of the domain has been established in this way, step 5 is the task of iterative encoding of the ontology through logical formalization. This is achieved by transforming the natural language definitions for the terms contained in the ontology into a format that is computer usable using an ontology editing tool.

While the process of five steps is top down in nature, working from the highly general to successively less general terms in the ontology, in practice it will involve a great deal of cycling via feedback loops between the successive steps. In the following sections, we will discuss the demarcation and information-gathering processes in more detail. In chapter 4 we will address the issue of regimentation, and we will return to issues of encoding in chapter 8.

Explicitly Determine the Subject Matter of the Domain Ontology

The first step in constructing a domain ontology is to determine explicitly the intended scope of the ontology, which is to answer the question: "what part of reality is this ontology an ontology of?" Providing an explicit statement of this scope will serve to indicate both what is to be included in and what is to be excluded from the intended ontology. For example, the documentation for the Foundational Model of Anatomy (FMA), an ontology of human anatomy, describes the ontology as "strictly constrained to 'pure' anatomy, i.e., the structural organization of the body."[1] This statement makes it clear which terms are candidates for inclusion in the FMA, but also which terms to exclude: those relating, for example, to functional anatomy, or surgical anatomy. The specification of scope will also indicate the level or levels of granularity of reality against which the ontology is calibrated. Will this be *multicellular organisms*, or *organs,* or *cells*, or *cell components*, or *molecules*? Or will it perhaps be *entire populations of organisms*? Or will it be some combination of levels, as in an ontology that deals with cell signaling and thus needs to represent, for example, both cells and signaling pathways?

Domain and Top-Level Ontologies

We have seen that for purposes of successfully managing scientific information in the long term, the root node or nodes of a domain ontology should be defined in terms of some highly general universal taken from a domain neutral ontology such as BFO. This will help to ensure that the ontology is built using a high-level ontology architecture that is shared with other ontologies.

If, for example, the relation *part_of* is asserted in a given formal ontology as being transitive (so that if *x part_of y* and *y part_of z*, then *x part_of z* will hold), then in a domain ontology built on its basis, for example, in the domain of anatomy, we will be able to use this feature of the parthood relation to infer from *finger part_of arm* and *arm part_of body*, to *finger part_of body*.

Similarly, if a top-level ontology contains distinct representations for continuants (three-dimensional entities that continue to exist through time, such as planets and molecules) and occurrents (entities that occur, which means that they are spread out not just in space but also in time, such as a baseball game or the movement of a planet in its orbit), then all domain ontologies defined on its basis will be required to respect this same distinction among the entities it represents.

In these and a series of related respects the top-level ontology helps to ensure the correctness of ontology construction at lower levels. If an ontology uses *part_of* but contains assertions that contradict transitivity, then these assertions can be flagged as being in need of manual checking. If an ontology recognizes the distinction between things and processes, then problem cases—for example, terms such as "gene mutation," which are ambiguous as between thing and process readings—can be identified in advance and warnings issued requiring developers to subject such terms to additional manual inspection. BFO was designed to play this kind of role in the process of domain ontology design and quality assurance.[2]

For these reasons it is important, at the outset of designing a domain-specific ontology, to consider what top-level ontological categories and relations might apply to the domain at hand, and to select a top-level ontology representing sufficient and sufficiently clear categories and relations to handle the basic kinds of entities to be found in the domain in question. It is important to note that, by definition, a top-level ontology should be domain neutral. Thus it should not contain representations of relations and universals that are specific to any given domain. It will thus, in comparison with many of the domain ontologies defined in its terms, be very small. The ontological content pertinent to each specific domain is then added to that of the top-level ontology through the process of downward population.

Relevance

The task of determining what portion of reality a domain ontology should represent involves also addressing the problem of determining what and how much of the existing data and information about a given domain should be included in the ontology. It can be summarized as the problem of determining what is *relevant* for the ontology, a matter to be determined (1) by the current state of science, and thus by the structure of the corresponding portion of reality, (2) by the degree to which existing ontologies in neighboring areas can be relied upon in supporting the given ontology's development, and (3) by the practical goals the ontology needs to satisfy.

What is objectively relevant to the Cell Ontology (CL), for example, is determined by the nature of cells themselves, what they are, what processes they characteristically initiate or are involved in, and so forth. The CL taxonomy of immune cells is created on the basis of information about the protein molecules expressed on cell surfaces; representations of relevant types of molecules are drawn from the Protein Ontology (PRO), to create definitions such as the following:

lymphocyte of B lineage = def. lymphocyte and (has_plasma_membrane_part some CD19 molecule) and (lacks_plasma_membrane_part some CD3 epsilon)

Or in other words: a lymphocyte of B lineage is a lymphocyte that has CD19 molecules on its plasma membrane but does not have CD3 molecules on its plasma membrane. Here "lymphocyte" is a higher-level term defined in CL, "CD19 molecule" and "CD3 molecule" are defined in PRO, and "plasma membrane" is defined in the Cellular Component branch of the Gene Ontology.

Connections between cells and proteins are handled by building links between relevant ontologies in such a way that the information compiled in each of these ontologies is brought together in ways useful for reasoning and integration. By this means we avoid also some of the dangers of silo formation—for example, where those interested in cells feel tempted to develop their own local ontology of protein surface markers, an ontology that would fail to interoperate with other protein information resources. Ensuring that the corresponding domain ontologies have been structured on the basis of the same top-level ontology from the start makes it easier to bring them into alignment in the needed way.

The task of determining what needs to be represented in an ontology will also depend on the practical goals the ontology needs to satisfy. All ontology development (like all of science) is to some degree opportunistic: which parts of an ontology are developed first, or in the greatest detail, will often depend on available funding, and such funding will often be bound to purpose. Goal-oriented human activities bring some entities into focus and leave others in the background. If our job is to support the scientific investigation of a hypothesis relating to, say, fetal diseases involving lymphocytes of B lineage, then we will first identify existing ontologies with relevant content. But our investigation may require the development of an entirely new ontology focusing strictly on specific areas—for instance on interactions between these and those specific kinds of cells and molecules in these and those kinds of patients undergoing these and those kinds of treatment.

These ways in which purpose can determine ontology content reflect the distinction introduced in chapter 2 between *reference ontologies* and *application ontologies*. A reference ontology is a representational artifact analogous to a scientific theory, in which maximizing expressive completeness and adequacy to the facts of reality are of primary importance. An application ontology is a representational artifact that is designed to assist in the achievement of some specific goal. Reference ontologies will be constructed and structured primarily on the basis of the established content of a scientific discipline. Application ontologies will be constructed and structured primarily in terms of what is relevant to some specific goal. To be successful in the long term, however, application ontologies should to the maximal possible extent make use of portions of reference ontologies as their starting points. Development of application ontologies thereby can also benefit work on reference ontologies, for example, where terms created within the framework of the former are discovered to be of general scientific relevance, for example, then these terms will be promoted to a level where they form part of a reference ontology available for more general use.

Granularity

One subpart of the problem of determining relevance is the issue of determining the appropriate granularity of the entities to be represented in an ontology. The issue of granularity arises because things in reality, and also the parts of things, come in many different sizes and possess varying degrees of complexity. There is a continuum stretching from subatomic particles, atoms, and molecules, through ordinary objects such as animals, rocks, and tables, on to ecosystems, planets, solar systems, galaxies, and ultimately the universe itself. There is an analogous continuum also in the realm of processes unfolding in time, stretching from milliseconds to years to geological epochs. Things and processes can be identified at all of these different granularity levels, and as we move up to successively coarser grains we encounter entities that exhibit features not found at lower levels—a phenomenon referred to by philosophers under the heading of "emergence." The problem of granularity for ontology design is the problem of deciding the prototypical sizes and complexities of entities that are to be represented in a given domain ontology. Should an ontology of mountains include representations of the types of molecules of which the mountains are composed? Should an ontology of stages in the plant life cycle include stages of growth of individual leaves? When setting out to design an ontology, the choice of root nodes will determine, in part, the level or levels of granularity that will form part of the ontology's coverage, but this determination will be influenced primarily by the needs of users of the ontology—for example, in reflection of the degree to which finer gradations of taxonomy enable the recording of differences in data of a practically useful sort.

The Problem of Nonexistents

Once the domain or scope of an ontology has been decided on, what is needed is a systematic survey of the content of the established science relevant to this domain. This means primarily surveying the current content of authoritative textbooks and of salient terminologies. Thus ontologies relate primarily to the use of general terms in established sciences. Ontologies may in areas such as chemistry be used to represent types of entities which do not exist—for example, molecules not yet synthesized—but in general the rule is that ontologies should consist of representations only of those types for which we have good evidence that instances exist (and, by extension, only of those defined classes for which we have good evidence that there are members).

Very occasionally, ontologies may need to be developed to support research in areas still subject to dispute among different groups of scientists and thus not belonging to established science. (Recall, again, the case of the "Higgs boson.") We prefer to see such ontologies as provisional in nature, to be promoted to the ranks of ontologies proper only when the disputes in question have been settled. The methods for the creation of such provisional ontologies will then be essentially the same as those outlined here, but will apply the process of term selection not to established textbooks but, for example, to journal articles produced by some subset of the disputing partners. The results of such provisional ontology development will then also be provisional. They will be able to be added to existing ontology content and treated like other ontologies only when the disputes in question have been resolved.

Conclusion

In this chapter we have introduced some general principles of ontology design and provided an overview of the two initial phases of the ontology building process—namely, *demarcating the domain of the ontology* and *gathering information about the domain*. In the following chapter, we will discuss the third step of the ontology building process: regimentation, which deals in greater detail with issues of terminology selection, definition, and classification.

Further Reading on Relevance, Perspectivalism, Granularity, and Adequatism

Bittner, Thomas, and Barry Smith. "A Theory of Granular Partitions." In *Applied Ontology: An Introduction*, ed. Katherine Munn and Barry Smith, 125–158. Frankfurt: Ontos Verlag, 2008.

Hill, David P., Barry Smith, Monica S. McAndrews-Hill, and Judith A. Blake. "Gene Ontology Annotations: What They Mean and Where They Come From." *BMC Bioinformatics* 9 (suppl. 5) (2008): S2.

Kumar, Anand, Barry Smith, and Daniel Novotny. "Biomedical Informatics and Granularity." *Functional and Comparative Genomics* 5 (2004): 501–508.

Masci, Anna M., Cecilia N. Arighi, Alexander D. Diehl, Anne E. Lieberman, Chris Mungall, Richard H. Scheuermann, Barry Smith, and Lindsay G. Cowell. "An Improved Ontological

Representation of Dendritic Cells as a Paradigm for all Cell Types." *BMC Bioinformatics* 10 (February 2009): 70.

Rector, Alan, Jeremy Rogers, and Thomas Bittner. "Granularity, Scale and Collectivity: When Size Does and Does Not Matter." *Journal of Biomedical Informatics* 39 (2006): 333–349.

Smith, Barry. "Ontology (Science)." In *Ontology in Information Systems: Proceedings of the Fifth International Conference* (FOIS 2008), ed. C. Eschenbach and M. Gruninger, 21–35. Amsterdam: IOS Press, 2008.

Smith, Barry, and Werner Ceusters. "Ontological Realism: A Methodology for Coordinated Evolution of Scientific Ontologies." *Applied Ontology* 5 (3–4) (2010): 139–188.

Smith, Barry, et al. "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration." *Nature Biotechnology* 25 (11) (November 2007): 1251–1255.