Inflammation Ontology Design Pattern: an exercise in building a core biomedical ontology with Descriptions and Situations

Aldo GANGEMI[a], **Carola CATENACCI**[a], **Massimo BATTAGLIA**[b]

[a]*Laboratory for Applied Ontology, ISTC-CNR, Rome, Italy*
[b]*IMS-CNR, Rome, Italy*

Formal ontology has proved to be an extremely useful tool for negotiating intended meaning, for building explicit, formal data sheets, and for the discovery of novel views on existing data structures. This paper describes an example of application of formal ontological methods to the creation of biomedical ontologies. Addressed here is the ambiguous notion of inflammation, which spans across multiple linguistic meanings, multiple layers of reality, and multiple details of granularity. We use UML class diagrams, description logics, and the DOLCE foundational ontology, augmented with the Description and Situation theory, in order to provide the representational and ontological primitives that are necessary for the development of detailed, flexible, and functional biomedical ontologies. An ontology design pattern is proposed as a modelling template for inflammations.

1. Introduction

Ontology is the talk of the day in most information science domains. Ontological artifacts are able to solve many problems, which are characterised as special *semantic services*.

For example, suppose that a healthcare unit from a hospital *a* needs to get information about a patient from a patient record which has been encoded in a different hospital *b*, and that it has to check the updatability of that information according to the system used in the hospital *a* in order to provide the necessary (and minimal) information to the units that carry on tests (so that they can schedule them), to the administrative unit (so that it can manage the patient's case), to the patient herself (so that she receives a readable but precise résumé of her case), etc.

A scenario of the kind sketched above involves a lot of heterogeneity in the information that enters the flow, as well as in the information that is produced by the flow. Several services are involved, multiple formats, multiple database schemata, more than one communication protocol, etc. Up to now, the only strategy that promises to produce a unified layer for reasoning about all those aspects is their reformulation according to an *explicit semantics*.

A semantic service can be defined as an information service that is provided according to an explicit *formal semantics*. A formal semantics provides a logical foundation to the languages used in the specification of services and their domains of interest (e.g. those exemplified in the previous paragraph). Roughly speaking, when a language has a formal semantics, the representation of a content in that language is a *formal theory*. In their most appropriate form, ontologies are formal theories.

For example, the database schema from hospital *b* (as well as other possible schemata) is a formal theory that can be assumed as an ontology, but also the communications delivered to units, professionals, and patients can be generated out of (or be based upon) ontologies.

Different ontologies can have complex interrelationships when are used together to carry out composed services. The resulting structure is called here *conceptual architecture*.

The Semantic Web is maybe the most apparent application area for semantic services [19,28], but the services that are addressed by the Semantic Web are not fundamentally different from those traditionally dealt with by database integration, requirement analysis, conceptual modelling, information integration, meaning negotiation, etc.

The current innovation leverages the availability of languages and theories which can (or must) be shared and reused beyond a *local* agreement. Therefore, the *global* nature of semantic services should be considered in its wide range of implications.

The Semantic Web, but even much narrower complex structures (e.g. the one envisaged in the opening scenario) are very prone to generate an "ontological chaos".

A conceptual architecture is required because of the main use of ontologies: making intended meaning available to all (artificial or human) agents that could be involved in a semantic service. Intended meaning is bound to the *context* in which expressions of a language are used, such as physical situations, theoretical frameworks, social norms, plans and goals, linguistic practices, etc. Hence, the representation of intended meaning needs a very flexible and rich set of primitives that can be put within a *modular* architecture, either across ontologies or across elements of an ontology. The intended result of this approach is the design of ontologies which, whatever the task they are meant to accomplish, are of a *high quality*, and thus able to avoid the ontological chaos that could arise from an undervaluing of conceptual architecture.

Biomedicine is a typically very complex domain as far as intended meaning is concerned: different activity domains (e.g. clinical vs. administrative knowledge), different scientific granularities (e.g. molecular vs. organic detail), different user requirements for the same service (e.g. physician-oriented vs. patient-oriented views), historically ambiguous terminology (lexical polysemy), etc. Moreover, biomedical knowledge is a unique blend of competences and data types that can be used to improve or create semantic services, e.g. (besides the classical information integration example reported above) testing hypotheses on distributed bioinformatics data, enhancing educational information, verifying the consistency of guidelines, enabling evidence-based medicine, etc.

Conceptual architectures can be seen as explicit representations of context dependencies. Such dependencies are not bound to formal languages or theories. For example, services such as information retrieval and multiple database access might use different degrees of expressivity of ontology specification languages (e.g. OWL-Lite vs. OWL-DL [19]) and of ontological theories (e.g. a taxonomy vs. a conceptual model of inflammations), but, in principle, they could require analogous conceptual architecture complexity. When we try to represent a domain of interest by means of ontologies, we have to deal, in most cases, with heterogeneous knowledge and with different levels of detail of that knowledge. This complexity is not bound to logical expressivity: whether we want to retrieve documents or to find complex knowledge patterns in distributed databases, we need to develop an ontology library with an accurately designed internal stratification of modules and concepts. For example, we may want to distinguish:

- A biologist's - vs. a healthcare manager's - intended meaning of "inflammation"
- Different explanation levels for an inflammation condition (e.g. biochemical vs. organic)
- A description of diagnostic constraints vs. a set of patient data
- Different *roles* played by the same entity, e.g. a chemical operating both as a drug and as an aetiology

An explicit conceptual architecture requires a *strongly modular* ontology design. Strong modularization refers to the capability of *talking about contexts*, or – formally – of having contexts in the domain of discorse of a theory or set of theories, *together with* the elements of those contexts [24,15]. Unfortunately, existing modularization methods are inadequate, either because they lack a proper semantics (e.g. XML-like *namespaces*), or because they are confined to *soft modularization*, as with partitions of ontologies or knowledge bases (e.g. the so-called *microtheories* [1], or *formal contexts* [2]).

In this paper, we present an exercise in applying a strongly modular architecture to some biomedical notions related to *inflammation*. Firstly, we describe the problem addressed in the exercise, then we introduce the methods and tools used in the exercise, and, finally, we present the results, in the form of a core ontology of inflammations. In the conclusions, some related issues and future work are mentioned.

2. The problem and its dimensions

In the first place, we should note that biomedical ontologies to be used in a semantic service can be entrenched in complex ways. For example, interactions between clinical and biological knowledge require more than two domain ontologies: they need to be supported by a more general, *core* biomedical ontology. Moreover, different existing core biological and biomedical ontologies [3,4,5,6,7,8,9,23,25] may require a more general, *foundational* ontology in order to interoperate.

This fundamental interaction can be supported by a typical *soft modularization* of ontologies into strata and modules (Fig. 1), which induces certain choices at the top stratum: which is the foundational ontology that implements the most appropriate set of principles [15], i.e. the most proper *vision* of the world? A typical example of "vision" is the one encoded in most Western natural languages and cultures, sometimes called the *3D view* of the world: objects are distinct from processes, attributes are predicated of either objects or processes, space is primitive, etc. (for alternative visions, see [15]).

Soft modularization induces *specificity* choices at lower strata, for examples by separating general, *core* notions of a domain from more specific ones, or by singling out typical *patterns* of conceptualization that are central of a domain and can constitute a single module in the core stratum.

The conceptual architecture provided by soft modularization is an essential step in developing high quality ontologies. On the other hand, inducing a vision and a certain specificity detail is not enough, because - even within the same vision and with the same set of core and domain ontologies - different views can be considered by domain experts or can be required by semantic services to be deployed.

The intended meaning of some biomedical expressions (e.g. *inflammation*), for example, could be explicitly bound to an intended *context* (e.g. a clinical condition), or could refer directly to a context

(e.g. to the diagnostic guidelines applied to that condition) or to some role played within a context (e.g. to the process model, the value types, or the typical objects that can be found in a condition).

In order to talk of these contextual notions, we need a *strong modularization*, which is hard-wired into the ontology itself. In other words, in our domain of discourse we want both *things* and the *way of doing things*. How to represent a way of doing things in an ontology? This problem has not been addressed yet in ontological engineering, although some authors appear to be partly aware of both the theoretical and practical sides of the problem [10,11][1]. Recently, some work carried on at our laboratory has started facing the problem [17,24], and some results are already used in several application domains.

The notion of *inflammation* illustrates these issues neatly. We will introduce the notion from three different but related points of view: *linguistic polysemy*, *conceptual detail*, and *epistemological layering*.

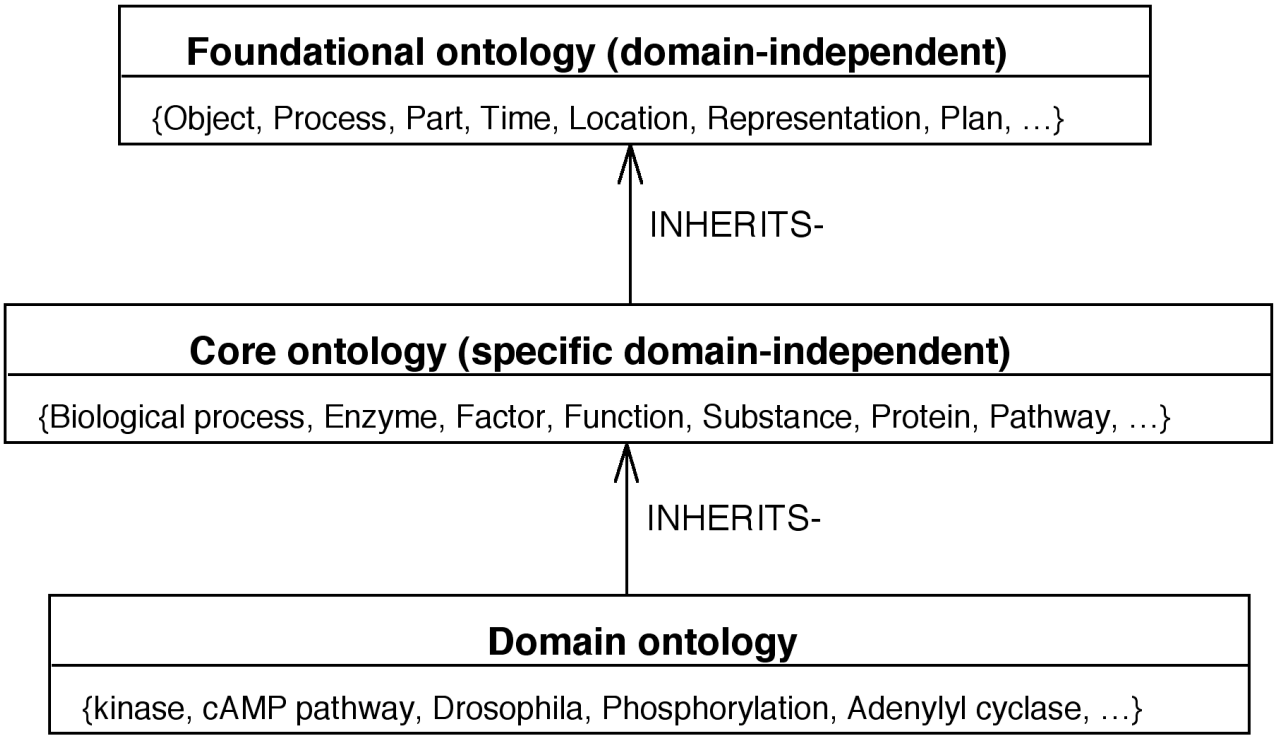

Figure 1. Ontology library stratification, with an example of biochemical concepts

2.1 Linguistic polysemy

Polysemy is the linguistic phenomenon by which a term has more than one meaning. It is widespread both in everyday and specialized languages, and medicine makes no exception. Far from being a negligible issue in information management, it may seriously hinder crucial services such as intelligent information access, natural language processing or terminological standards definition.

[1] Outside ontological engineering, much work has been done in AI, e.g. in planning, situation calculus, process modelling, etc. This research, however, has concentrated mainly on algebras of actions and processes rather than on an integrated ontological characterization of actions, plans, and objects. Moreover, most frameworks require some second-order-logic constructs.

Since polysemy is a natural linguistic phenomenon, it cannot be ignored in a truly semantic web, which has to rely on solid conceptual foundations for an effective and unambiguous terminology management.
"Inflammation" is a typical polysemous word. At least five different meanings of inflammation can be recognized in the different contexts in which the term appears [5], i.e.:

1. Inflammation *segregates* external agents
2. The inflammation has a *diameter* of 5 cm.
3. The inflammation has changed its *shape*
4. The inflammation evolved *during* three weeks
5. The inflammation is *severe*

This evidence has been reported as natural for domain experts, but conceptually *hard* for ontologies (e.g., in the GALEN project [3]), and this has been proposed as a case for (or against) foundational ontologies: are they an acceptable framework for the description of that evidence? In particular, the senses 1 to 5 may be represented in existing biomedical taxonomies or thesauri as follows:

1. A *physiological function* performing segregation of external agents, since segregation seems to be a *consequence*, and consequences usually hold between events (e.g., in SNOMED [6], or in UMLS semantic network [7])
2. A characteristic portion of a *body part*, which embodies that physiological function. In fact, a *diameter* should be an attribute of an object rather than one of an event
3. A specific abnormal *morphology* of that portion, since morphology is the category for shapes (e.g. in SNOMED)
4. A clinical *condition* encompassing all those entities: conditions *evolve*, hence they seem to concern situations rather than single events or objects
5. A *diagnosis* applicable to that condition, since *severe* appears to be a *qualification* given in the context of an *assessment* rather than an attribute of an event or objective condition as such.

In our opinion, instead of being a tricky challenge, this evidence constitutes a rich source of information for a methodology based on foundational ontologies, because of the hints it gives us on the *underlying core* ontology of inflammation, in the form of a *polysemy network* [12]. In order to let the core ontology emerge, we should commit on a foundational ontology that contains an adequate set of primitives (concepts and relations), with enough axioms to capture the relations hidden in the polysemy network. Such relations are *cognitively present*, as anyone can realize by reading the previous list of senses and possible taxonomic categorizations. Our job is to give to these relations an explicit and rigorous systematization.

2.2 Conceptual detail

Before introducing the ontologies that we have reused in order to build a core ontology of inflammation, we should also mention other sources of ambiguity, which are less obvious than those provided by linguistic evidence.

A second source of ambiguity, at least for the agents that may want to use biomedical ontologies for providing sophisticated services, is the *explanation detail* of biomedical notions. In the inflammation case, for example, we can list the following detailed events as typical of an inflammation condition:

- Response to a lesion (stimulus) causing cellular death or suffering
- Elimination of causal agent
- Removal of cellular debris
- Repair of damaged tissue

Is each of these events an inflammation? Are they *parts* of an inflammation? And in which of the senses listed above?

Several other lists can be provided that constitute plausible accounts of inflammation, e.g. the classical list of "symptoms" given by Celsus in the I century a.C, and integrated by Virchow in the XIX century: *rubor, calor, tumor, dolor; functio laesa*.

An example of deepening detail level can be retrieved e.g. from a recent educational overview [13]:

- Alteration in microvasculature
    - Changes in vascular flow and diameter
        - Transient vasoconstriction of arterioles

Further detailed views can be found if we focus on the so-called "goals" of inflammation:

- Circumscribe lesion
- Eliminate cause
- Repair tissues in order to recover function

or on *agents*:

- Exogenous (infection, trauma, temperature changing, toxins)
- Endogenous (ischemia, neoplasm, autoimmune pathology)

or on *participants* in inflammatory response:

- Vascular tissue (endothelial cells of vessels, blood cells, granulocytes, monocytes, lymphocytes, plasm proteins)
- Connective tissue (mast cells, fibroblasts, proteins of extracellular matrix)

or on *attributes*:

- Acute vs. Chronic
- Type of vascular response
- Amount of infiltrated cellular substance (physical region)
- Type of lesion agent
- Duration (temporal region)
- Signs/symptoms (possible constituents of an inflammation condition, with a different epistemological status, e.g. *pain* vs. *WBC count*)

The explanation detail of the intended meaning of "inflammation", as it is available to experts, varies from gross distinctions to fine accounts of processes, objects, and attributes. Relying on bottom-up input is important, but it does not suffice: we need methodological and conceptual tools that enable us to move from a detail level to another, without loosing the dependencies that hold across different levels.

2.3 Epistemological layering

The need for ontological consistency with reference to a common framework becomes particularly clear when we take into account the phenomenon of *explanation layering*. By "layering" we mean the structure of reality that is assumed to exist in a scientific (or even common sense) domain for a certain descriptive purpose. For example, "inflammation" can be viewed either as a *systemic*, an *organic*, a *cellular*, or a *molecular* concept. These layers are partly independent within the overall conceptual organization of biomedical research, but when getting to grips with a clinical phenomenon, they reveal their dependencies, as the current development of bioinformatics and molecular medicine is increasingly showing.

An example of layering for inflammation knowledge is the following [13]:

- Acute inflammation (*systemic layer*)
    - Accumulation of leukocytes at site of tissue injury (*organic layer*)
        - Adhesion by interactions btw families of cell molecules (*cellular layer*)
            - ELAM-1 synthesized by cytokine-activated cells (*molecular layer*)

The same source also reports an inflammation biochemical pathway (Fig. 2) involving:

1) The process by which antibodies are complexed by antigens
2) The formation of enzymes (convertase)
3) The binding of enzyme molecules to antigenic surface and complexed antibodies, and finally
4) The "inflammation" induced by the *breakdown products* from previous steps in the pathway.

This pathway assumes a layering to explain the emergence of an inflammation condition from biochemical processes. The inflammation pathway also shows a kind of ordering, which can be interpreted either temporally or causally.

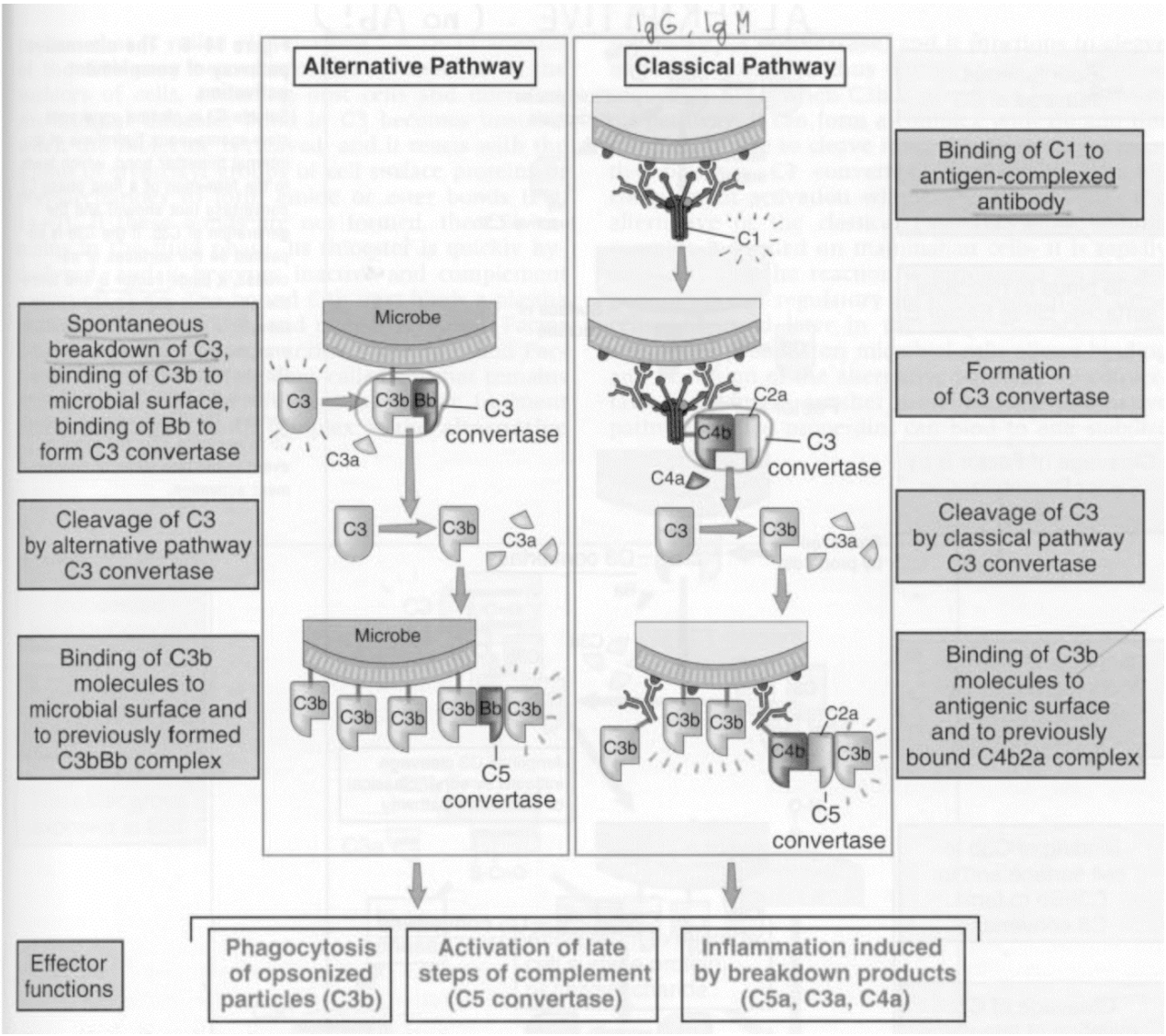

Figure 2. Inflammation biochemical pathway

3. Methods and tools

In this exercise, we will assume the following frameworks: the ONIONS [5] and ONIONS-II [14] methodologies for terminology analysis and ontology integration and merging, the DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) foundational ontology [15] and its extensions (http://www.loa-cnr.it), specially including the ontology of *Descriptions and Situations* (D&S) [15,17,24], and the WonderWeb methodology for ontology quality assessment [16]. We will not analyse here the pros and cons of these frameworks compared to other proposals, since we would like to concentrate on the general approach, which is more or less shared by the designers of other foundational ontologies and formal methodologies. Incidentally, DOLCE is only the first of several modules that populate the "WonderWeb Foundational Ontology Library" under development in the homonymous project (http://wonderweb.semanticweb.org).

3.1 Some methodological principles

The main requirements and criteria adopted in our formal ontology methodologies can be summarized as follows:

Logical consistency. Ontologies should be expressed in a logical language with an explicit formal semantics, e.g. FOL, KIF, any DL[18], OWL[19], etc.

Groundedness of primitives. The primitive classes and relations used in an ontology should be *grounded* in some set of principles that reflect e.g. cognitive motivation (implying intersubjectivity and large reusability) or empirical appropriateness (implying local soundness). Grounding is maybe the most difficult requirement to fulfil, since such fulfilment relies on evidence coming from external domains such as philosophy, cognitive sciences, and linguistics.

Semantic coverage. An ontology should assume to cover all the entities of its domain (*extensional* coverage), and it should also try to cover all the entity *types* of its domain (*intensional* coverage). For example, the assumption that all enzymes involved in inflammation pathways are in our domain of interpretation concerns extensional coverage, while a commitment to all the enzyme types that (according to the state of our knowledge) are conceivable in our domain of interest is an example of intensional coverage.

Modelling precision. An ontology should try to represent all and only the *intended models* for its domain of interest and for the tasks the ontology is meant to accomplish. For example, an ontology of biochemical inflammation pathways would be *overcommitted* if it tried to represent relationships between enzymes and administrative healthcare concepts; it would be *undercommitted* if it failed to represent relationships between enzymes and the substances that play an enzymatic role.

Soft and strong modularity. Ontologies should modularize the domain's conceptual space as much as possible, by organizing the domain theories either into different ontologies or into different *descriptions* and *situations*. For example, "inflammation" as a healthcare concept will belong to a healthcare ontology module, while as a biochemical concept it will belong to a biochemical ontology module. Moreover, inflammation as a diagnostic concept (a description) will be different from inflammation as a condition (situation). In the terms of the D&S theory, inflammation as a condition *satisfies* inflammation as a diagnosis (see section 3.2).

Scalability. Expressive languages should be used to represent detailed accounts of intended meanings, according to the domain and tasks to be accomplished. Expressive ontologies can be used either as *reference* ontologies or to support sophisticated services, such as multiple databases querying. As reference ontologies, they can be scaled in order to support computationally hard services, e.g. information extraction from large repositories or retrieval across the Semantic Web. The scaled-down versions are "lightweight" versions of reference ontologies. Reference versions can be maintained independently, and can be used to produce updates of the lightweight versions according to knowledge management schedules.

Ontology design patterns. The most general notions that are defined in visions (foundational ontologies) and in specific domains (core ontologies) are often richly interrelated, usually in a way that allows domain experts to reuse the same set of relations for most of their modelling work. These highly interconnected fragments of foundational or core ontologies can be called *ontology design patterns* (ODP) [26], analogously to design patterns used in software engineering [27]. The research on the compositional properties of ODPs, and on their validity as best practices in

ontology development has just begun, but experiences with users seem encouraging. These are some desiderata for ODPs:

- An ODP is a template for solving a domain modelling problem
- An ODP "extracts" a fragment of a foundational or core ontology, which is its "background"
- An ODP is axiomatized according to the fragment it extracts
- An ODP can be represented in any ontology representation language, although its intuitive and compact *visualization* seems an essential requirement
- An ODP can be an element in a partial order, where the ordering relation requires that at least one of the classes or relations in the ODP are specialized
- An ODP can be intuitively exemplified and catches relevant core notions of domains
- An ODP can be often built from informal or simplified schemes used by domain experts, together with the support of a foundational ontology and a methodology for domain ontology analysis, or by specializing existing ODPs
- An ODP can/should be used to describe a best practice of modelling
- An ODP is similar to a DB schema, but an ODP is defined with reference to a foundational ontology and should have a general character, independently from local design details.

3.2 Reused ontologies

In our exercise, we have reused DOLCE-Lite-Plus (DOLCE+), which is an extension of the DOLCE foundational ontology [15] with the Descriptions and Situations (D&S) ontology [17], in order to have enough primitives to represent the problem dimensions.

For the sake of visual clarity, here we represent our ontologies as non-standard UML class diagrams, assuming a description logic-like semantics [18] for them: *classes* are interpreted as *concepts*, *generalization* is interpreted as *formal subsumption* (subclassOf relation), *associations* and *attributes* are both interpreted as *binary relations*, except that associations allow also to display the cardinality of inverse relations.[2]. Where no cardinality is indicated, it is 0..*. We also assume an access to a reasoning service that supports concept classification and value maps (or relation chaining). The ontologies mentioned here are available in various languages and format (cf. http://www.loa-cnr.it/ontologies).

[2] Various semantics have been proposed for UML (e.g. [35]). Here we use class diagrams as a diagrammatic interface to the description logic ("lite") version of DOLCE-Lite-Plus.

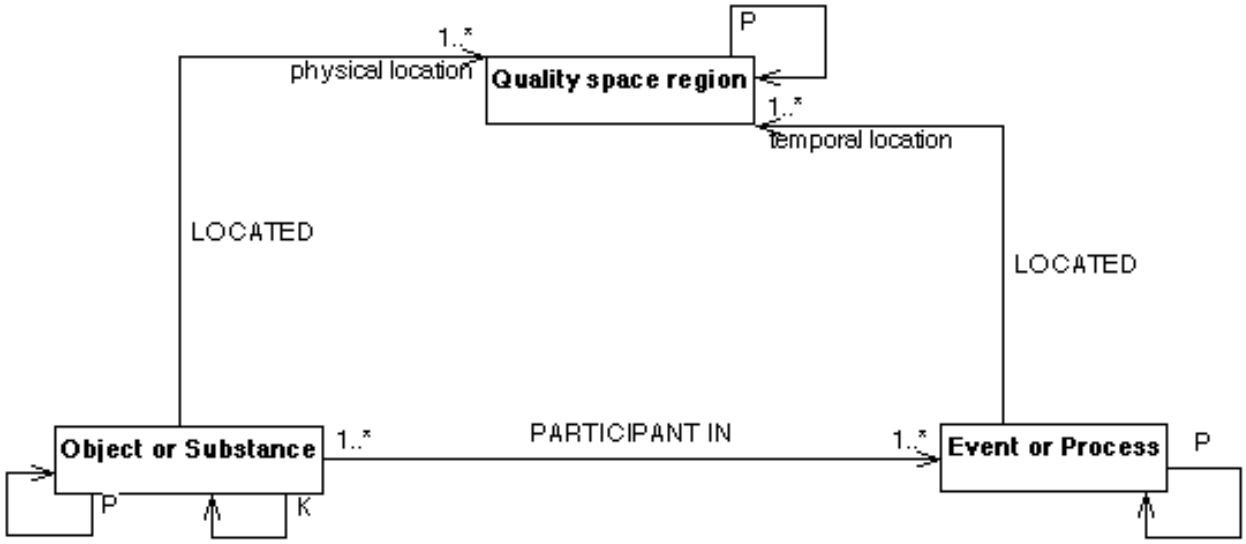

Figure 3. The basic design pattern derivable from DOLCE. P means *part*, K means *constitution*. This pattern can be instantiated to say basic things such as «object *o* (with its parts and constituents) is located in place *s* and participates in (some or all of the parts of) the event *e*, which is located at time *t*». By means of the background DOLCE axioms, various inferences can be drawn from these basic facts.

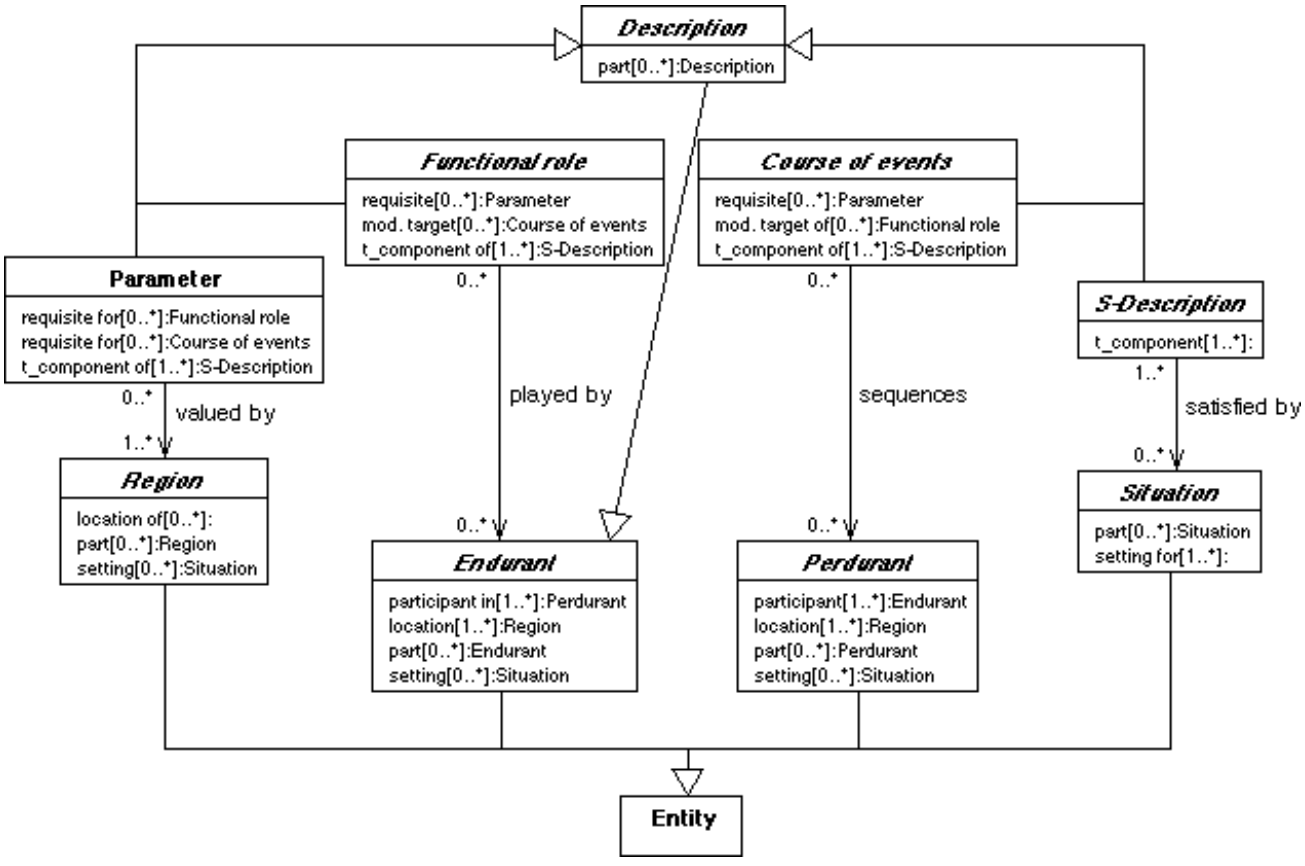

Figure 4. The Descriptions and Situations (D&S) design pattern.

In Figure 3, a basic design pattern that derives from a central fragment of DOLCE is shown. Figure 4 displays a more complex design pattern derived from the central fragment of the D&S extension of DOLCE. Due to their flexibility in representing the core concepts and relations in many different domains, these ontology design patterns are currently being used in several academic and industrial projects (see http://www.loa-cnr.it/projects).

Pure DOLCE has four basic top categories: *endurant* (including object- and substance-like entities, either physical or not), *perdurant* (process-, event- and state-like entities), *quality* (individual attributes), and *abstract* (including mainly conceptual *regions* that are approximately attributes); and several primitive relations, such as *part*, *connection*, *constituency*, *inherence* of qualities in entities, *participation* of endurants in perdurants, etc.

The basic design pattern derivable from DOLCE allows to express systematic relations between objects, events, and attributes (Fig.3).

A more detailed set of primitives, however, is needed to express a clinically relevant ontology, which involves decision making entities, conditions, roles, etc. A good example of the types of entities required by clinical knowledge modelling is the original model of the Pen&Pad system for the Electronic Patient Record [29]. In Pen&Pad, an EPR is a *text* representing the *history* of a patient's *condition* (the "observation" of the latter), the related *decision making* process, and the *information attributes* of itself or of other sources. In other words, an EPR ontology needs a domain of discourse which includes theories (diagnostic contexts), guidelines (diagnostic process models), and information objects (texts). Hence, we need to talk about *knowledge objects*.

The Descriptions and Situations ontology (D&S) supports a first-order manipulation of *knowledge* (or *descriptive*) *objects* (such as plans, diagnoses, norms, institutions, etc. – i.e., *theories*) and *situations* (such as cases, facts, settings, etc. – i.e, *models* or *states of affairs*), as well as it allows a characterisation of the elements of descriptions and situations. In practice, D&S is an ontology of *rationality*. In principle, it is adaptable to any foundational ontology, but it is here demonstrated as an extension to DOLCE.

D&S domain includes reified theories, concepts, and states of affairs:

- Reified theories are called *s-descriptions* (or ***contexts***)
- Reified concepts (theory elements) are called *c-descriptions* (or ***roles***)
- Reified states of affairs are called *situations* (or ***configurations***)
- An s-description can be *satisfied by* a situation
- An s-description is *composed by* c-descriptions
- A situation is *constituted by* entities in the ground ontology (in this case, DOLCE)
- A c-description has systematic relations with sibling c-descriptions, and with entities in the ground ontology (see below)
- *States of affairs* can be extracted from the world in many ways
- *Situations* reify states of affairs from the viewpoint of an s-description
- A situation must satisfy at least one *s-description*
- This means that no state of affairs is reified in D&S unless it has an explicit s-description

These (here informally described) axioms implements the so-called *constructivist stance*: world views are constructed by the intentionality of an observer *o*, although depending also on entities that exist independently from *o* (for related assumptions, see e.g. [20,21,22,31,32]). This means that the ontological commitment laying behind any application of D&S depends entirely on what one is dealing with (i.e., the conceptualizations embedded in the domain), and can vary accordingly. A practical consequence of this stance is that providing the combinatorial preprocessing of all possible states of affairs is useless, because descriptions provide the identity criteria for those that are relevant.

Many uses of D&S involve matching an independently introduced situation to existing s-descriptions (e.g. clinical conditions vs. diagnoses). In fact, a situation can be matched either against the description it depends on (e.g. in plan execution assessment), or against a new description (e.g. in clinical diagnoses). The second case is what we may call a *redescription*.

These mechanisms are inspired by cognitive phenomena like Gestalt psychology *figure-ground shifting* [31] and *representational change* in cognitive development [32].

C-descriptions reify *constraints* on which classes can be types of the constituents of a situation:

- The c-descriptions that reify a constraint on DOLCE regions (e.g. "blood pressure" or "body temperature") are called *parameters* (e.g., "critical systolic blood pressure" or "fever"). For example, notice that *fever* "selects" a sub-region of "body temperature".
- The c-descriptions that reify a functional property of DOLCE endurants (e.g. "enzyme" or "person") are called (*functional*) *roles* (e.g. "mediator" or "patient"). For example, notice that *patient* "selects" a functional property of "person".
- The c-descriptions that reify abstract sequences of DOLCE perdurants (e.g. "therapeutic procedure" or "biochemical process") are called *courses* (e.g. schedules, algorithms, or pathways). For example, notice that *pathway* "selects" a particular course of biochemical processes.

Some specific relations hold among c-descriptions: *modality-target* holds between functional roles and courses, and *requisite-for* holds between parameters and either functional roles or courses. *Modality-target* reifies the modal dependence between a functional property and a sequence, e.g. a physician may be *allowed* or *disallowed to* administer a certain therapy, or an enzyme might be *possibly* or *certainly* involved in a step of a pathway.

Situation constituents and c-descriptions are systematically related. The basic relation, *selects*, reifies the instantiation relation between an individual in a model and a concept in a theory. When applied to DOLCE, "selects" relates c-description to instances of DOLCE categories. Intuitively, *selects*(*x,y*) binds an individual *y* classified in a DOLCE category to a situation *s* that *satisfies* the s-description *d* that has *x* as a component. In particular: parameters are *valued by* regions, functional roles are *played by* endurants, and courses *sequence* perdurants (Fig. 4).

An example of a D&S application is the following: a clinical condition (situation) has an associated diagnosis (s-description) made by some agent (object) participating in the diagnostic procedure (situation), in which the agent plays a role (physician), according to some diagnostic guideline (s-description) that has a sequence of steps (course), and certain thresholds (parameters) as components, etc.

4. Some results

4.1 An ontology design pattern for inflammation descriptions

According to the strata depicted in Fig. 1, an ontology of inflammation would be an instance of a *core* ontology. As such, it reuses the DOLCE+ foundational ontology, and can be used by a domain ontology, e.g. of specific inflammations or other clinically relevant domains.

In Fig. 5, the D&S schema from DOLCE+ is applied to inflammations[3]. The linguistic evidence for the ambiguity of "inflammation" presented in section 2 can now be revisited:

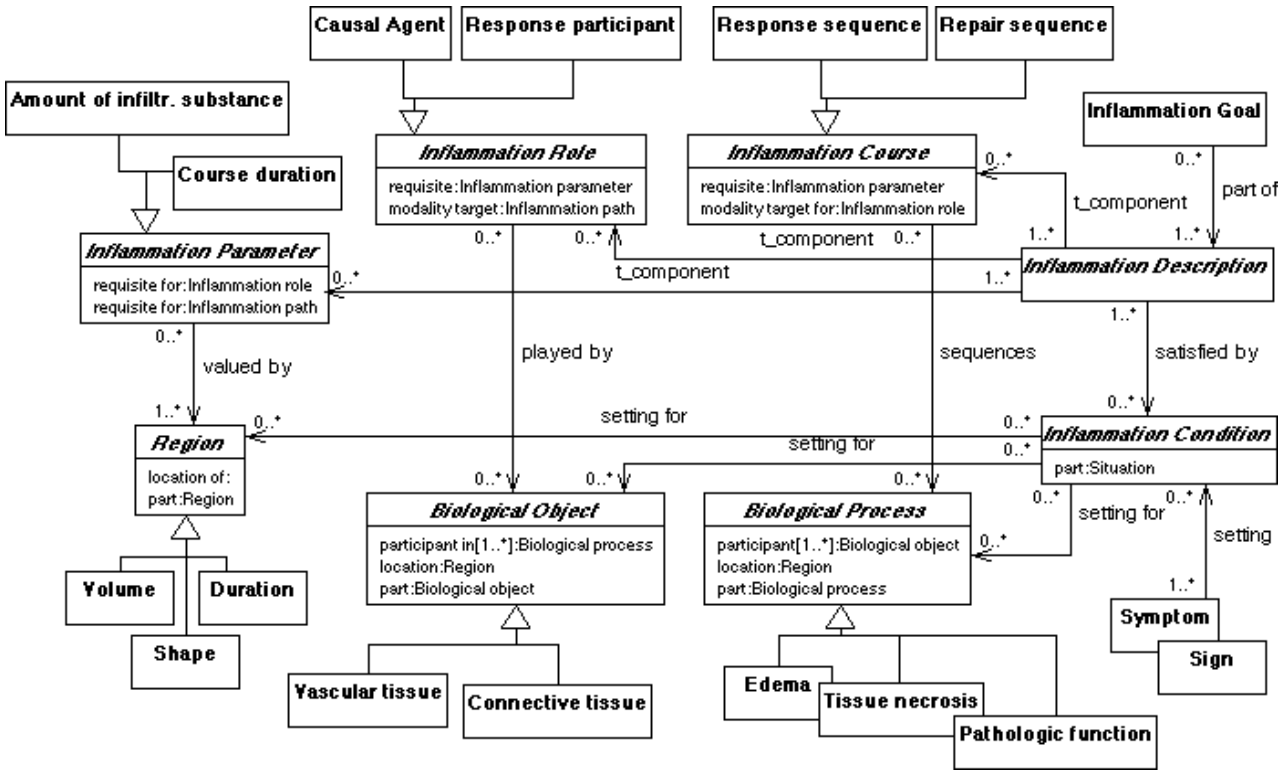

Figure 5. The inflammation design pattern. This pattern specializes the D&S pattern, and its classes subsume the core classes involved in the description of inflammations (symptom, connective tissue, repair sequence, causal agent, etc.). Moreover, it shows the systematic polysemy network related to the term *inflammation*.

1) Inflammation as a *physiological function* (*inflammation#1*) is a kind of biological process (a kind of perdurant)
2) Inflammation as a *characteristic portion of a body part* (*inflammation#2*) is a feature (a relevant part of an object) that embodies *inflammation#1*

[3] This schema, along with the discussion on the ambiguity of "inflammation", has already been presented in [33].

3) Inflammation as a *specific abnormal morphology* (*inflammation#3*) is a quality (represented by a region in a space of values) inherent in *inflammation#2*

4) Inflammation as a *clinical condition* (*inflammation#4)* is a situation, i.e. the setting for the entities (*inflammation#1, #2, #3*, etc.) whose relations have been evidenced through a diagnostic procedure or some other assessment.

5) Inflammation as a *diagnosis* (*inflammation#5*) is an s-description satisfied by *inflammation#4* iff *inflammation#4* constituents (including also *#1, #2, #3)* are compatible with its components (courses, functional roles, and parameters).

"Inflammation as a diagnosis" is called in our ontology "inflammation description". A diagnostic description is adequate to some clinical condition ("inflammation condition") only if the constraints posed by the components of the diagnostic description are satisfied by the condition's elements.

C-descriptions are called here: 1) "inflammation course", which "sequences" "inflammation as a process" (the biological processes typical of the inflammation condition); 2) "inflammation role", which can be "played" by "inflammation as inflamed object"(the biological objects involved in the condition); and 3) "inflammation parameter", which can be "valued by" by "inflammation as morphology" (a kind of region, i.e. the attributes related to the inflamed objects).

4.2 Epistemological layering in the inflammation design pattern

The D&S for inflammations can represent data structures that satisfy constraints defined within *layered* contexts (see section 2.3). For example, we have seen that explanations of inflammation might span from the molecular and organic layers to the systemic, clinical one. How can these layers be considered in a same ontological theory?

We can show the dependencies among such epistemological layers (Fig. 6) through an ontology in which a situation that satisfies an s-description at a lower layer (e.g. *molecular*) can be a *part* of a situation that satisfies an s-description at a higher layer (e.g. *organic*), which in turn can be a part of a situation at an even higher level (e.g. *clinical*). For example, a biochemical pathway of inflammation will be satisfied by certain biochemical situations that can be part of the systemic situations that satisfy an organic description of inflammation. Systemic situations can be part of the condition that satisfies a clinical diagnosis.

The reason why we can stack situations in this mereological structure actually relies on the possibility of *embedding* an s-description within a situation at a lower layer. For example, the systemic description of an inflammation *plays a role* defined within the clinical description, and the molecular description *plays another role* defined within the systemic description.

Consequently, mereological stacking of situations depends on role-embedding of s-descriptions (and possibly their components). This feature of D&S allows additional expressivity that mirrors the typical cognitive stacking used by humans for finding the minimal constraints necessary to carry out some task [34].

Notice that a soft modularization of ontologies for the different layers would not allow this kind of reasoning, unless many specialized bridging axioms are provided according to a meta-level theory.

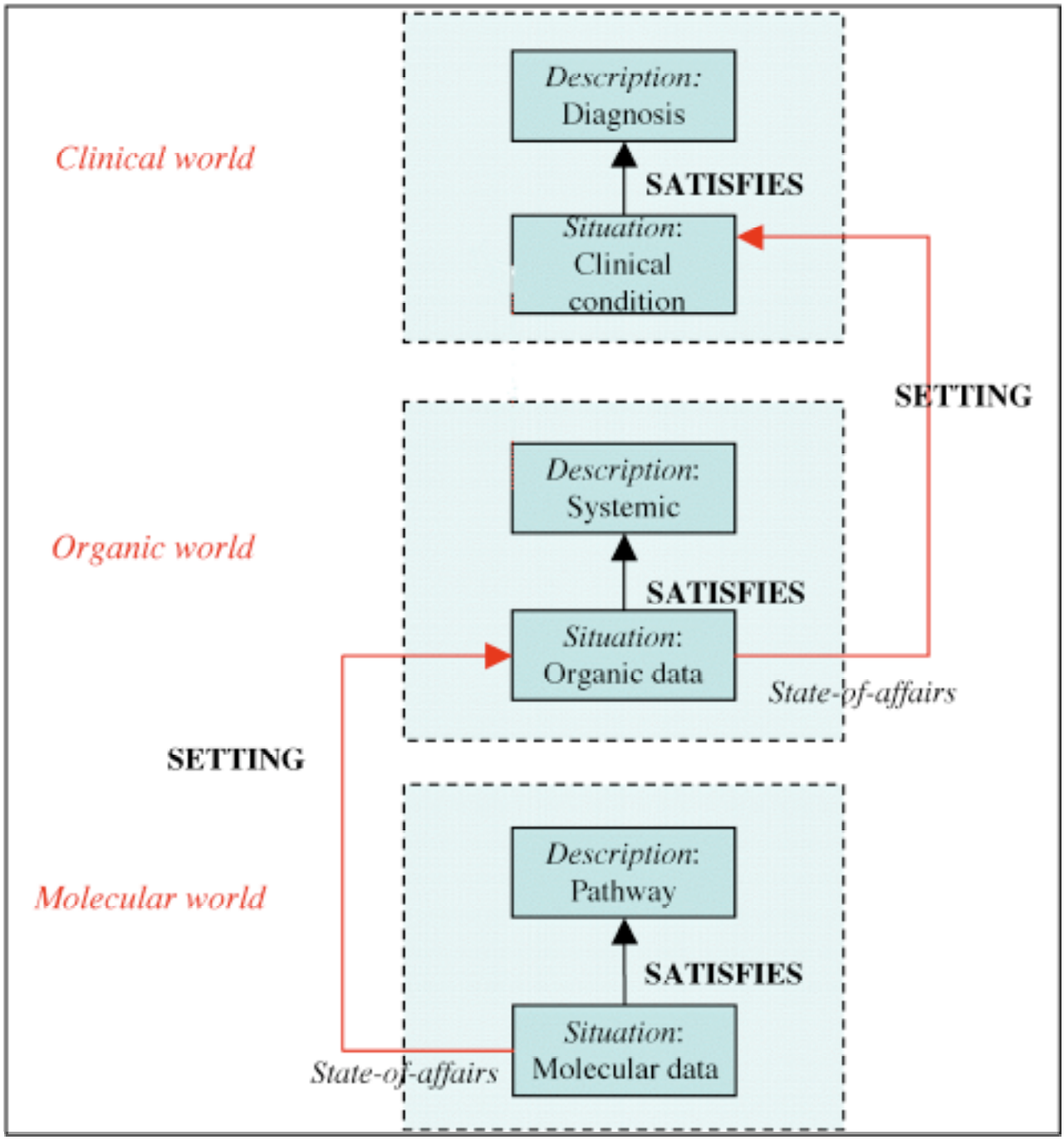

Figure 6. A diagram showing an epistemological layering of inflammation-related knowledge. An inflammation biochemical pathway description can be satisfied by a molecular data structure (thus reified as a situation under the pathway description). The molecular data structure then becomes a constituent of an organic data structure, which is reified as a situation under an organic inflammation description. The organic data structure becomes, in turn, a constituent of a clinical data structure, which is reified as a clinical condition under a diagnostic inflammation description. Constituent ("setting") relations holding between the clinical condition and the lower situations (organic and molecular) are inferable via the background axioms of D&S. Other axioms provide embedding of lower descriptions into the higher ones (see text).

5. Conclusions

We have presented an exercise in core domain ontology development. Axiomatic theories, grounded primitives, and ontology design patterns have been employed to characterize the complex notion of *inflammation* across different related conceptualizations, explanation details, and epistemological layers.

The ontology of inflammation so designed provides an analogue of the cognitive context available to an expert. For example, this ontology, together with other suitable tools for information extraction, might enable a semantic service to contextualize any particular occurrence of *inflammation* with a degree of accuracy comparable to that of a human expert.

This exercise in ontological engineering shows how much flexibility is required to ontological engineering in order to provide exhaustive content to (semi-)automatic biomedical reasoning. How much will we be able, in the future, to diagnose an individual against an inflammation, relying on both biochemical features and systemic observation? We are convinced that foundational and core ontologies should be part of the scenario. They can serve as a unique ontological framework that provides the primitives to represent different contexts, explanation layers and detail levels, along with all their mutual relationships.

More generally, we have described in a semi-formal way[4] a quite detailed exercise that exhibits the development of core domain ontologies and ontology design patterns, i.e. the reusable components for domain ontology building. This work and the resulting components are *complementary* to semi-automatic ontology learning techniques aiming at building large-scale, poorly axiomatized, possibly low quality ontologies. In fact, although reusable components are primarily needed for heavy reasoning services, they are ideally suited for the refinement of ontologies that have been generated through machine learning procedures (see e.g. [5,30]).

An evaluation of the results obtained in this exercise should be based on a multi-dimensional test suite, which takes into account at least: a) domain expert acceptance, b) degree of "meaning negotiation" allowed by the ontology, c) efficiency of semantic services implemented on top of the ontology, d) degree of "intensional coverage" compared to that resulting from the barely logical integration of existing heterogeneous sources (e.g. thesauri), e) efficacy in refining ontologies produced through machine learning techniques from texts, f) degree of scalability of the ontology without loosing core functionalities. Such a test suite has not been attempted yet in any domain, and we are willing to build it in the near future.

References

1. RV Guha, "Contexts: A Formalization and Some Applications", PhD thesis, Stanford University (1991).
2. F Giunchiglia, C Ghidini, "Local Models Semantics, or Contextual Reasoning = Locality + Compatibility", *Artificial Intelligence*,127(2):221–259 (2001).
3. A Rector, J Rogers, "Ontological Issues in using a Description Logic to Represent Medical Concepts: Experience from GALEN", IMIA WG6 Workshop: Terminology and Natural Language in Medicine (1999).
4. I Papakin, B Smith, K Munn, W Ceusters, "Ontology and Bodily Systems", AMIA2003 (2003).
5. A Gangemi, DM Pisanelli, G Steve, "An overview of the ONIONS project: Applying ontologies to the integration of medical terminologies", *Data and Knowledge Engineering*, 31(2):183–220 (1999).
6. KA Spackman, KE Campbell, RA Coté, "SNOMED RT: A reference terminology for health care", Proceedings of AMIA 97 Conference (1997).

[4] The complete axiomatization would take much more space and is possibly outside of the scope of this book; incidentally, it can be downloaded separately from the internal report repository of our laboratory.

7. A McCray, “An Upper-Level Ontology for the Biomedical Domain”, *Journal of Functional Genomics*, II (2003).
8. A Burgun, O Bodenreider, “Aspects of the taxonomic relation in the biomedical domain”. In: C Welty C. Second international conference on formal ontologies in information systems, Ogunquit, Maine (2001)
9. DM Pisanelli, A Gangemi, G Steve, “The Role of Ontologies for an Effective and Unambiguous Dissemination of Clinical Guidelines”, in R Dieng, O Corby (eds.), Knowledge Engineering and Knowledge Management. Methods, Models, and Tools, Berlin, Springer-Verlag, pp. 129-139 (2000).
10. JF Sowa, “Conceptual Structures: Logical, Linguistic, and Computational Issues, chapter Ontology, Metadata, and Semiotics”, pages 55–81. Number 1867 in Lecture Notes in AI. Springer Verlag, Berlin (2000).
11. A Ankolekar, M Burstein, J Hobbs, O Lassila, D Martin, D McDermott, S McIlraith, S Narayanan, M Paolucci, T Payne, K Sycara, “DAML-S: Web Service Description for the Semantic Web”, *Proceedings of the International Semantic Web Conference (ISWC*) (2002).
12. A Gangemi, DM Pisanelli, G Steve, “Understanding Systematic Conceptual Structures in Polysemous Medical Terms”, in: J Marc Overhage (ed.), Proceedings of the 2000 AMIA Annual Symposium (2000).
13. E Adelstein, M Balas, “Acute Inflammation, Cellular Events, Chemical Mediators of Inflammation”, http://www.muhealth.org (2003).
14. A Gangemi, “Some Tools and Methodologies for Domain Ontology Building”, *Journal of Functional Genomics*, II (2003).
15. C Masolo, S Borgo, A Gangemi, N Guarino, A Oltramari, L Schneider. The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable 18, http://wonderweb.semanticweb.org (2003).
16. S Borgo, A Gangemi, N Guarino, C Masolo, A Oltramari. Ontology RoadMap: Ontology infra structure for the Semantic Web, WonderWeb Deliverable 15, http://wonderweb.semanticweb.org (2002).
17. A Gangemi and P Mika, “*Understanding the Semantic Web through Descriptions and Situations*”. Meersman R, et al. (eds.), Proceedings of the International Conference ODBASE03 (2003).
18. F Baader et al., “The description logic handbook”, Cambridge: Cambridge University Press (2003).
19. M Dean, D Connolly, F van Harmelen, J Hendler, I Horrocks, D McGuinness, PF Patel-Schneider, and LA Stein, “OWL Web Ontology Language 1.0 Reference”, Technical report, World Wide Web Consortium (W3C) (2002).
20. J Austin, “How to Do Things with Words”, Harvard UP (1962).
21. T Winograd, F Flores, “Understanding Computers and Cognition”, Addison-Wesley (1987).
22. JR Searle, “The Construction of Social Reality”, New York, Free Press (1995).
23. E Ratsch, “Developing a Protein Interactions Ontology”, *Journal of Functional Genomics*, II (2003).
24. C Masolo, L Vieu, E Bottazzi, C Catenacci, R Ferrario, A Gangemi, N Guarino, “Social Roles and their Descriptions”, C Welty and D Dubois (eds.), Proceedings of KR04, Springer (2004).
25. Baker, P.G., Goble, C.A., Bechhofer, S., Paton, N.W., Stevens, R. and Brass, A. An Ontology for Bioinformatics Applications, Bioinformatics, Vol 15, No. 6, 510--520 (1999).
26. A Gangemi, “Ontology Design Patterns”, Technical report 2004#1, Laboratory for Applied Ontology (http://www.loa-cnr.it) (2004).
27. E Gamma et al., “Elements of reusable OO software”, Addison-Wesley (1995).
28. D Oberle, P Mika, A Gangemi, M Sabou, “Foundations for service ontologies. Aligning OWL-S to DOLCE”, Proceedings of WWW2004, Springer (2004).
29. AL Rector, WA Nowlan and S Kay, “Unifying Medical Information using an Architecture Based on Descriptions”, *Published in RA Miller (ed) Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care,* IEEE Computer Society Press (1990).
30. A Gangemi, R Navigli, P Velardi, “The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet”, Meersman R, et al. (eds.), Proceedings of ODBASE03 Conference, Springer (2003).
31. W Köhler, “Gestalt Psychology”, Liveright, New York (1947/1929).
32. A Karmiloff-Smith, “Beyond Modularity: A Developmental Perspective on Cognitive Science”, Cambridge, MA.: MIT Press (1992).
33. D M Pisanelli, A Gangemi, M Battaglia, C Catenacci, **“**Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies”, Proceedings of MEDINFO 2004, San Francisco (2004).
34. A Karmiloff-Smith, “Constraints on representational change: Evidence from children’s drawing”, Cognition, 34: 57-83 (1990).
35. G Guizzardi, H Herre, G Wagner, "Towards Ontological Foundations for UML Conceptual Models", 1st International Conference on Ontologies, Databases and Application of Semantics (ODBASE'02), Springer-Verlag, Berlin (2002).