THE ONTOLOGY OF BLOOD PRESSURE: A CASE STUDY IN CREATING ONTOLOGICAL PARTITIONS IN BIOMEDICINE

A. KUMAR[a], B. SMITH[b,c]
[a]Laboratory of Medical Informatics, Dept. of Computer Science, University of Pavia, Italy.
[b]Institute for Formal Ontology and Medical Science, University of Leipzig, Germany.
[C]Department of Philiosophy, University at Buffalo, New York, USA.
anand@aim.unipv.it; phismith@buffalo.edu

We provide a methodology for the creation of ontological partitions in biomedicine and we test the methodology via an application to the phenomenon of blood pressure. An ontology of blood pressure must do justice to the complex networks of intersecting pathways in the organism by which blood pressure is regulated. To this end it must deal not only with the anatomical structures and physiological processes involved in such regulation but also with the relations between these at different levels of granularity. For this purpose our ontology offers a variety of distinct partitions – of substances, processes and functions – and integrates these together within a single framework via transitive networks of *part-whole* and *dependence* relations among the entities in each of these categories. The paper concludes with a comparison of this methodology with the approaches of GO™, KEGG, DIP and BIND and provides an outline of how the methodology is currently being applied in the field of biomedical database integration.

1 BASIC FORMAL-ONTOLOGICAL DISTINCTIONS

Continuants, sometimes called endurants, are entities which *endure*, or *continue to exist*, through time. They are entities which preserve their identity from one moment to the next. Examples are: organisms, blood, the heart, the circulatory system. Occurrents, in contrast, are entities which *occur* in given intervals of time. They never exist in full in any single instant. Rather, they unfold themselves through time in the way in which the process of blood circulating through the body unfolds itself in time (Bittner & Smith, 2003).

To say that an entity is *dependent* is to assert that it requires a support from other entities in order to be sustained in existence. There is no *Cardiac output* without a heart and no *Cellular motion* without some cell which moves. *Independent* entities, in contrast, require no support of this sort from other entities. They are the substrates for entities in other categories.

Dependence relations can obtain also between dependent entities themselves and we can represent the networks of such relations in graph-theoretical terms (Smith, 1992). At some stage, though, dependence relations must bottom out, in entities which are not themselves dependent on anything else. Two important subspecies of dependence relations linking dependent entities can be distinguished.

First is the relation of *realization*, which binds a function (or role, or power) to a process when the former is realized (exercised, performed, expressed) in the latter. Second is the relation of *functional dependence*, which connects those occurrent entities whose values vary in tandem with each other. *Cardiac output* is functionally dependent in this sense on the processes taking place inside the arteriole and postcapillary vessels. Note that *Cardiac output* is functionally dependent also on other processes, for example on those involved in physical exertion, emotional crisis, blood transfusion, injury and change in ambient temperature. For reasons of space, however, the modifications associated with the latter are not treated here.

Because occurrents in the domain of biomedicine are always changes or movements *of* or *in* some enduring entity or entities, they are always dependent entities. Thus, of the four possible combinations yielded by the two divisions of continuant/occurrent and dependent/independent entities, only three are instantiated, namely: dependent continuant, independent continuant, and occurrent.

Dependent continuants include *functions* (for example the function of the heart: *to pump blood*), *states*, *qualities* and *roles*. The *exercise* of a function, in contrast – like the *performance* of a role, the *realization* of a disposition, and *changes* in a state or quality – is a *process*, an occurrent. We shall use the term 'activity' to designate those processes which arise when functions are exercised.

2 THEORY OF GRANULAR PARTITIONS

When human beings seek to classify the entities in a domain such as that of blood pressure, then they *partition* those entities into *cells* (Bittner & Smith, 2003a). Different partitions may represent cuts at different levels of granularity. Each partition consists of cells and subcells, the latter being nested within the former in a hierarchical manner. A given partition can consist of many layers of cells and subcells (for example in the animal kingdom the layers of genus, species, family, order, phylum, kingdom, etc.). Some simple conditions on partitions can be stated as follows.

A1: Every partition has a unique maximal or root cell in which all other cells are included as subcells.

A2: The subcell relation is reflexive, antisymmetric, and transitive.

A3: Each cell in a partition is connected to the root via a finite chain of immediate succeeding cells.

A4: If two cells within a partition overlap, then one is a subcell of the other.

These axioms ensure that each granular partition can be represented as a tree in the graph-theoretical sense. For a fixed domain that is partitioned at different levels of granularity, entities apprehended in partitions of finer granularity are *parts* of entities apprehended in partitions of coarser granularity: *membranes* are parts of *cells*; *cells* are parts of *organs*; *organs* are parts of *organisms*; and so on.

<u>*Navigating between partitions.*</u> In constructing an ontology of complex domains in biomedicine, a variety of partitions are needed, no one of which can be complete.

Each considers a given domain from within a certain context, placing some entities in the *foreground* and others in the *background* (Smith & Brogaard, 2003). Moving from one domain to another at the same level of granularity – for example from *Regulation of blood pressure* to *Regulation of renal function* – is a case of *ontological regrouping*; switching from a partition of some fixed domain at one level of granularity to a partition of the same domain at some other level of granularity a case of *ontological zooming*.

3 PHYSIOLOGY OF BLOOD PRESSURE

Blood pressure is the pressure exerted on the arterial walls by the flow of blood. The term 'blood pressure' designates both a certain *function* (to *exert* pressure) and also a certain *state* (of pressure being exerted), both of which are dependent continuants, the latter being associated with a *value*. The process of change in this value is an occurrent, which is dependent on the underlying state. Like all other functions in the body, blood pressure is subject to *regulation*. Regulation of blood pressure involves the exercise of a number of different functions in different parts of the body. Their collective task is to maintain blood pressure value within a certain interval.

Blood pressure is clearly related to cardiac output (the function expressed by the heart in pumping blood) and to peripheral vascular resistance (the resistance of the passage of blood through the precapillary arterioles). Blood pressure values are maintained within the relevant range by moment-to-moment regulation of cardiac output and of peripheral vascular resistance exerted primarily at the level of the arterioles, postcapillary venules and heart (Katzung, 2000). The most important dimensions of this regulation are as follows:

– The *heart* contributes to the maintenance of blood pressure via cardiac output (that is: by pumping blood).

– The *kidney* contributes by regulating the volume of the fluid present in the blood vessels (intravascular fluid), thereby controlling the power of the vessels to transport blood.

– *The internal cellular lining of the walls of the blood vessels* regulates vascular resistance via local release of hormones such as endothlin-1 and nitric oxide.

– The *baroreceptors* (sensory nerve endings in the walls of the auricles of the heart, aortic arch and carotid sinus) are responsible for the rapid moment-to-moment adjustments in blood pressure affected by postural changes. Central sympathetic neurons arising from the vasomotor area of the medulla are perpetually active in maintaining the tone of the blood vessel walls. Baroreflexes are stimulated by the stretching of the blood vessels brought about by the pressure of the blood. These act in combination with *the renin-angiotensin-aldosterone system* and other mechanisms related to body fluids to coordinate the functions of the mentioned vascular sites.

Where baroreceptors are responsible for short-term blood pressure control, the kidneys, by controlling the volume of blood passing through the vessels, are pri-

marily responsible for long-term control. A reduction in the pressure of the vessels supplying blood causes a decrease in bloodflow to the kidneys and a concomitant increase in their absorption of salt and water. In addition, decreased pressure in the renal arterioles as well as sympathetic neural activity via beta-adrenoceptors stimulates production of the hormone renin, which leads in turn to an increase in production of the hormone angiotensin II. The latter then causes direct constriction of the blood vessels and stimulates the synthesis of the hormone aldosterone in the adrenal cortex. This increases sodium absorption by the kidneys, which increases the osmotic pressure of the intravascular fluid and thus increases intravascular blood volume.

An ontology of blood pressure must find a place for all of the entities – substances, functions, processes, states – involved in the above. These can be classified in two lists. First, we have the basic entities, the more important of which are represented in Table 1. Second, we have compounds derived from these, built up by using operators like 'Regulation of', which are discussed in the text to follow.

Table 1. Classification of basic entities in blood pressure ontology

Independent Continuant	**Dependent Continuant**	**Occurrent**
Adrenal cortex	Angiotension II function	Adrenal cortex activity
Aldosterone	Baroreceptor function	Aldosterone activity
Angiotension II	Baroreflex function	Angiotension II activity
Arterioles	Blood pressure	Arterioles activity
Arterial wall	Blood pressure value	Artery activity
Artery	Cardiac output	Autonomic nerves activity
Autonomic nerves	Central sympathetic neuron function	Baroreceptor activity
Baroreceptors	Endothelin-1 function	Change in blood pressure
Beta-adrenoceptors	Hormone function	Change in cardiac output
Blood	Intravascular blood volume	Central sympathetic neuron activity
Body fluids	Kidney function	Endothelin-1 activity
Central sympathetic neurons	Nitric oxide function	Hormone activity
Endothelin-1	Osmotic pressure	Kidney activity
Heart	Peripheral vascular resistance	Nitric oxide activity
Hormone	Postcapillary venules function	Change in peripheral vascular resistance
Intravascular fluid	Renin function	Postcapillary venules activity
Kidney	Renin-angiotensin-aldosterone system function	Salt reabsorption
Nitric oxide	Renal perfusion pressure	Sodium absorption
Postcapillary venules	Sodium function	Renin activity
Renin	Vascular endothelium function	Renin-angiotensin-aldosterone system activity

Renin-angiotensin-aldosterone system		Water reabsorption function		Sodium activity
Sodium				Vascular endothelium activity
Vascular endothelium				Water reabsorption activity

4 IMPLEMENTATION FORMALISM

Our ontology consists of interconnected partitions dealing with blood pressure, cardiac output and the associated renal functions. The machinery of ontological zooming and regrouping gives us the facility to navigate between these partitions and to bring to light connections between the entities foregrounded in each.

Partitions of the domain of substances, functions and processes based on the relation *part_of* form the principal axes of the ontology. For processes and functions, the boundaries are not always well-defined. Consider the functions *Regulation of blood pressure by cardiac output* and *Regulation of blood pressure*. Here it is not clear where the former ends and the latter starts. What is clear, however, is that any given case of *Regulation of blood pressure by cardiac output* is a case of *Regulation of blood pressure*. The former thus stands in a *part_of* relation to the latter.

Partition of regulatory functions concerning blood pressure This partition deals with those functions involved in blood pressure control which are exercised by determinate parts of the body and are of the form *regulation_of* (see Figure 1 below).

Parthood relations. The processes in the human body constitute pathways, in which the product of one reaction becomes the starting material for the next. All the pathways in the human body are regulated, whereby the regulators can be either substrates of processes within the pathway, or products of these processes, or external continuant entities. Regulation functions can be related together by parthood relations. Thus for example the parts of *Regulation of blood pressure* include:

Regulation of blood pressure by peripheral vascular resistance
Regulation of blood pressure by intravascular fluid volume
Regulation of blood pressure by arteriolar constriction
Regulation of blood pressure by cardiac output
Regulation of blood pressure by postural baroreflex
Regulation of blood pressure by renin-angiotensin-aldosterone system

Dependence relations. The dependence relations trace the dependencies among processes, functions and the corresponding parts of the human body. Thus the function *Regulation of blood pressure by cardiac output* is dependent on *Cardiac output*, a dependent continuant, which is dependent in turn on the *Heart*, *Arteriole* and *Postcapillary vessels*, all of which are independent continuants. Dependence relationships are transitive, which means that *Regulation of blood pressure by cardiac output* is itself dependent on the *Heart*, *Arteriole* and *Postcapillary vessels.*

In each such case we can define an ordering of immediate and mediate dependence in such a way as to generate a graph of the underlying dependence structure (Fine, 1995).

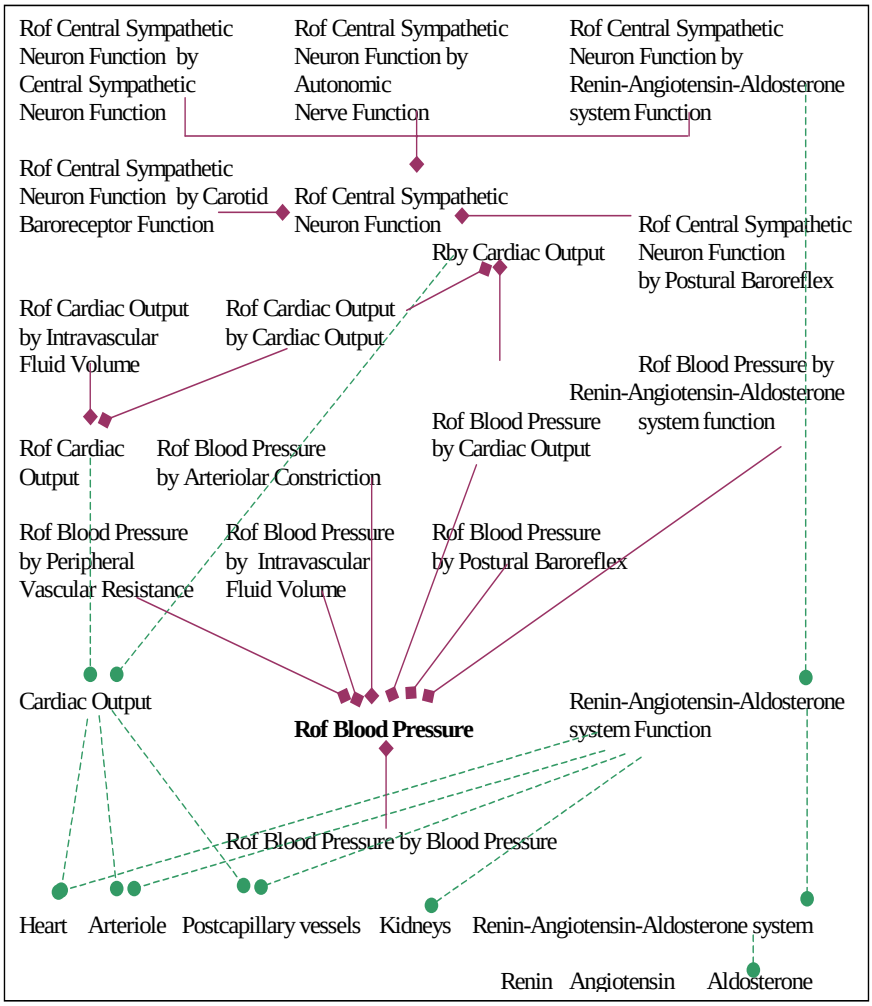

Figure 1. Sample partition related to *Regulation of blood pressure*
Rof = Regulation of; Rby = Regulation by; is-part-of; is-dependent-on

Ontological regrouping and zooming. Ontological regrouping occurs when for example we move from a partition at a given level of granularity related to *Regulation of blood pressure* to a partition at the same level of granularity related to *Regulation of renal functions*. Ontological zooming occurs when we move from a coarse-grained partition of, say, *Regulation of blood pressure* to partitions of this same domain at a finer grain. The coarser grained partition here would consist of cells foregrounding:

Regulation of blood pressure by cardiac output
Regulation of blood pressure by renin-angiotensin-aldosterone system function
Regulation of blood pressure by peripheral vascular resistance

The finer grained partition would foreground:

Regulation of blood pressure by beta-1 cardiac receptors function
Regulation of blood pressure by alpha-1 arteriolar receptors function

A still more fine-grained partition would relate to genetic level functions:

Regulation of blood pressure by 11 beta-hydroxylase gene function
Regulation of blood pressure by aldosterone synthase function
Regulation of blood pressure by renin gene function, and so on.

Ontological zooming can also involve moves within the category of dependent continuants, for example from *Regulation of cardiac output* to its lower-granular parts such as:

Regulation of cardiac output by intravascular fluid volume
Regulation of cardiac output by cardiac output.

The latter is an example of *autoregulation*, a phenomenon which occurs in almost every function and process in the human body and is represented graph-theoretically as a loop.

Regulation of renal functions and Regulation by blood pressure: The kidneys are related to the regulation of blood pressure in several ways, which include their regulation of the intravascular fluid volume, of renin-angiotensin-aldoesterone and of electrolytes. However, renal functions can also be partitioned in their own right and not merely as dependent upon (and as functioning in the service of) blood pressure regulation. Such functions include: *Regulation of glomerular filtration rate*, *Creatinine clearance*, *Clearance of sodium*. From a partition recognizing *Regulation by blood pressur*e on the part of the various functions depicted in Figure 1, we can zoom to a finer-grained partition recognizing *Regulation of glomerular filtration rate by blood pressure* as part of *Regulation of glomerular filtration rate* (see Figure 2). Furthermore, we can make the following ontological inference:

Regulation of glomerular filtration rate is-dependent-on *Glomerular filtration rate*,
Glomerular filtration rate is-dependent-on *Glomerular filtration*,
Glomerular filtration is-part-of *Renal function*.

From this we can infer that *Regulation of renal function by blood pressure* is-part-of *Regulation of renal function.*

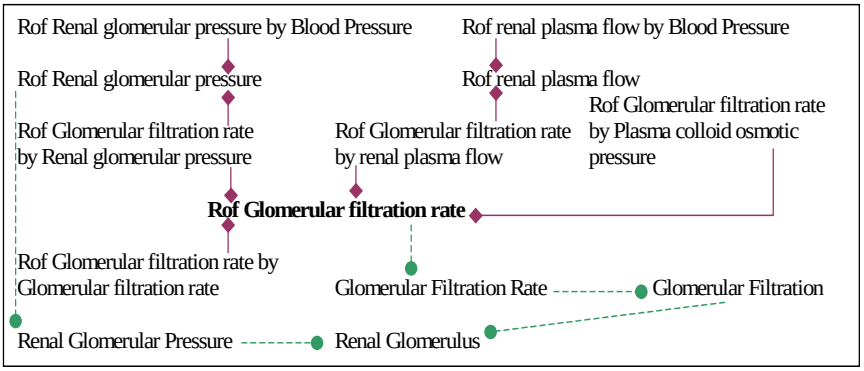

Figure 2. Sample partition related to Regulation of Glomerular filtration rate
Rof = Regulation of; Rby = Regulation by; is-part-of; is-dependent-on

Partition of processes related to blood pressure regulation Functions can exist even when they are not being exercised. Processes, in contrast, are such that, at any time at which they exist at all, they exist as actual realizations rather than as mere powers or potentials. Many (but not all) processes are indeed the *actualizations* of functions. Thus, if *Regulation of blood pressure* is a function, then the *Activity of regulating blood pressure* would be the related process.

Blood pressure regulation includes the heart's pumping of blood, the activity of the beta-1 receptors of the heart, peripheral vascular resistance, and the functioning of the alpha-1 receptors of the arterioles and of the aldosterone gene. *Alpha-1 arteriolar receptors* change their state just a few seconds after a change in *Blood pressure* (the change being mediated by activities involving catecholamines), while a change in *Aldosterone gene activity* will occur only many hours later. To do justice to these phenomena an explicit representation of temporal sequence is needed.

5 DISCUSSION

We can extract from the above the following requirements which, we believe, an adequate biomedical ontology should meet:

- ***R1***: Both *is_a* and *part_of* information should be explicitly represented. Parthood relations should be applied only between entities belonging to the same top-level formal-ontological categories.
- ***R2***: Networks of regulation of functions, including functions subject to autoregulation, should be explicitly represented.
- ***R3***: The ontology should trace both horizontal dependencies among dependent entities and vertical dependence relations to the underlying substrates – the

molecules, cells, organs and organ-systems – involved in regulating organic processes.

R4: The connections between the different substrates and processes in a pathway should be explicitly represented.

R5: Mediate and immediate dependencies between entities should be explictly represented.

R6: Levels of granularity should be explicitly represented, together with part-whole relations between entities on different granular levels.

R7: Functions and processes should be distinguished, and those processes which are the realizations of functions should be represented as such.

R8: An explicit representation of temporal sequence is needed.

We may now contrast the ontology proposed above with the treatment of biological functions and processes in familiar classification systems. Applying our methodology to the phenomenon of blood pressure has, we believe, thrown light on ways in which existing tools may need to be supplemented and we have taken first steps towards establishing potential areas of mutual benefit.

The Gene Ontology (GO™): GO (version: March 1, 2003) has played an important role in providing a controlled vocabulary for biological purposes. It provides a tripartite taxonomy under the headings: Cellular Components, Molecular Functions and Biological Processes. (Gene Ontology Consortium; http://www.geneontology.org/) The project GONG (for: Gene Ontology Next Generation) (Wroe *et al*, 2003) is attempting to improve GO's suitability for use in computers by rendering GO in a description logic.

GO has many strengths; but it also has certain weaknesses (Smith *et al*, 2003). Its application of *is-a* and *part-of* relations is problematic because the latter are applied between entities which belong to different formal-ontological categories (***R1***). Moreover, *is-a* and *part-of* are sometimes run togther. Thus GO has:

Regulation of blood pressure is-a *Circulation*
Circulation is-a *Physiological process*
Physiological process part-of *Biological process*

But *Regulation of blood pressure* is a function, and so if the *part-of* relation truly holds, then *Circulation* should be a function too. Then, however, it would no longer hold that *Circulation* is_a *Physiological process.* The second and third relations should thus be represented uniformly either as: *part-of* or: *is-a*. In addition, the part-hood relations between regulatory functions should be explicitly represented (***R2***).

GO is also in pertinent respects incomplete. Thus it has *Renin activity*, but not *Renin*. *Renin activity*, like all other GO activity terms, is currently represented as a daughter of *Molecular Function* (suggesting that, for consistency's sake, the latter node should be relabeled 'Molecular Activity').

GO is subject to weaknesses when dealing with substance terms. Thus it has:

Nucleus part-of *Intracellular*
Intracellular part-of *Cell*

Cell is-a *Cellular Component*

The fact that nucleus is not a part of all cells (for example, erythrocytes do not have a nucleus) is not itself a problem – this is in keeping with GO's reading of *part-of* as meaning: *sometimes-part-of*. GO has problems when it comes to dealing with time (for example with the opposition between long-term and short-term functioning). For even though it has nodes relating to processes and activities, the latter are still treated only within the context of statically conceived term-hierarchies. Thus GO has problems dealing with all those relations which hold during only a part of the life-cycle of a given entity. (***R7,8***).

Finally GO has problems when it comes to providing an account of how its three term-hierarchies are linked together, problems which are resolved in the framework presented above via the machinery of dependence relations.

Transparent Access to Multiple Bioinformatics Information Sources (TAMBIS): TAMBIS aims to provide integrated access to a range of biological databases (Baker, 1998; Baker 1999; http://imgproj.cs.man.ac.uk/tambis/). It includes a knowledge base of biological terminology, a model of the underlying data structure and a knowledge-driven user interface employing description logic. TAMBIS represents parthood relations via *isComponentOf* and functions via *hasFunction* and functional dependencies (***R1,3***). Our work is to this extent in alignment with TAMBIS. However, we go further by representing both part-whole relationships between functions themselves and also the dependence relations involved when functions are realized (***R5***). TAMBIS does not provide a detailed treatment of pathways, but it does provide a distinction between functions and processes (***R7,8***). It recognizes collective functions, but not autoregulation and not the functional organization generated by the two factors of *Regulation of* and *Regulation by* (***R2,3***).

Kyoto Encyclopedia of Genes and Genomes (KEGG): KEGG is focused on pathways. It integrates information pertaining to molecular interactions in biological processes with information about genes, proteins and the associated chemical compounds and reactions and provides dynamic links to various databases (Kanehisa *et al*, 2002; http://www.genome.ad.jp/kegg/kegg2.html). KEGG's LIGAND database contains information regarding compounds, reactions and enzymes, and allows searches of pathways which provide chains of events linking substrates to final products.

All interactions (including regulatory functions) are represented by means of a binary relation (***R2***). Functions and processes are not distinguished (***R7,8***). Moreover, the reciprocal influence of functions – including the regulation of a function with parts which in turn involve regulation by other functions – is not described (***R1,3***). Dependence relations can be inferred from existing links visible within the browser, but they are not explicitly represented within the system itself.

Database of Interacting Proteins (DIP) and *Biomolecular Interaction Network Database* (BIND): DIP integrates data regarding protein-protein interactions from different data sources (Xenarios *et al*, 2002; http://dip.doe-mbi.ucla.edu/). BIND is

a database which puts biological information together via three types of records – of interactions, molecular complexes, and pathways (Bader *et al*, 2003; http://www.bind.ca/). Neither of them are claimed to be ontologies by their respective authors; rather, they are curated collections of relations gleaned from the literature. However DIP and BIND would benefit from support for reasoning about part-whole relations and from the drawing of a distinction between functions and processes.

6 FUTURE WORK

It is recognized on all sides that tools are needed to allow dynamic navigation within heterogeneous databases and ontologies. Our approach provides one basis for the development of such tools, which is currently being implemented in a new type of ontology-assisted database integration framework (Verschelde *et al.*, 2003), which dynamically links a range of external heterogeneous databases to a central biomedical ontology. This dynamical linkage is supported by creating within the central ontology counterparts of the top-level concepts from the source databases in ways which allows the ontological relations already present in the central ontology to situate the information contained within the databases both in relation to the rest of the ontology and in relation to other linked databases.

The methodology has been applied by using GO as a reference for mapping protein data originating from the protein database Swissprot. The data and relationships in that database are mapped to ontology relationships within the central ontology, including the relationships of part-whole, dependence and regulation treated in the above. The resulting knowledge representation allows us to develop integrated views over heterogeneous data sources in ways which can serve as a starting point for data-mining and information extraction.

REFERENCES

1 Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* 31(1): 248-50, 2003.

2 Baker PG, Brass A, Bechhofer S, Goble C, Paton N, Stevens R. TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. An Overview. In *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, ISMB98, Montreal, 1998.

3 Baker PG, Goble CA., Bechhofer S, Paton NW, Stevens R, Brass A. An Ontology for Bioinformatics Applications, *Bioinformatics*, Vol 15, No. 6, 510-520, 1999.

4 Bittner T and Smith, B. Granular Spatio-Temporal Ontologies. *Proceedings of the AAAI Spring Symposium on Foundations and Applications of Spatio-Temporal Reasoning* (FASTR), 2003.

5 Bittner T and Smith B. A theory of granular partitions, *Foundations of Geographic Information Science*, M. Duckham *et al*, eds., London: Taylor & Francis, 117-151, 2003a.
6 Fine K. Part-Whole. In: *The Cambridge Companion to Husserl*. B. Smith and D. W. Smith (eds.), Cambridge: Cambridge University Press, 1995.
7 Gene Ontology Consortium. Creating the Gene Ontology Resource: Design and Implementation. *Genome Res*. 11: 1425-1433, 2001.
8 Katzung BG. *Basic and Clinical Pharmacology*. 8th edn. New York: Lange/McGraw-Hill. 155-157, 2000.
9 Kanehisa M, Goto S, Kawashima S, and Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res*. 30: 42-46, 2002.
10. Smith B and Mulligan K, Framework for formal ontology, *Topoi*, 3, 73-85, 1983.
12 Smith B and Brogaard B. A unified theory of truth and reference", *Loqique et Analyse*, 43: 169-170, 2000 (published 2003).
13 Smith B, Williams J and Schulze-Kremer S. The ontology of the Gene Ontology. *Proc AMIA Symp*. 2003.
14 Verschelde, J-L Santos M, Deray T, Smith B and Ceusters W. Ontology assisted database integration, *Proceedings of 2003 Bielefeld Symposium on Integrative Bioinformatics*.
15 Wroe CJ, Stevens R, Goble CA, Ashburner M. A methodology to migrate the gene ontology to a description logic environment using DAML+OIL. *Pac Symp Biocomput*. 2003: 624-35.
16 Xenarios I, Salwinski L, Duan XJ, Higney P, Kim S, Eisenberg D. DIP: The Database of Interacting Proteins. A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res*. 30: 303-5, 2002.

Acknowledgments: Work on this paper was supported by the Wolfgang Paul Program of the Alexander von Humboldt Foundation. We thank Olivier Bodenreider, Anita Burgun, Katleen Elsner and Dominikus Ernst for helpful comments on the manuscript.