Towards a Clinical Trial Ontology using a Concern-Oriented Approach

Crenguta Bogdan[1], Daniela Luzi[2], Fabrizio L. Ricci[3], Luca D. Serbanati[4]

[1] Ovidius University, Bd. Mamaia 124, 8700 Constanta, Romania
[2] CNR-IRPPS, Via Nizza 128, 00198 Roma, Italy
[3] CNR-ITB, Circ. Nomentana 496, 00162 Roma, Italy
[4] Politehnica University, Spl. Independentei 313, 5700 Bucharest, Romania

cbogdan@univ-ovidius.ro, {d.luzi, f.ricci}@irpps.cnr.it, luca@serbanati.com

Abstract. To reduce costs and improve the research quality in clinical trials (CT) a more systematic approach is needed for CT automatization as well as for enhancing interoperability at various levels of the research process. To this aim a conceptualization of the clinical trial should be carried out. Any modelling approach is based on partitioning criteria which enable us to dominate the complexity of the modelled universe. In this paper we introduce an original analysis method based on stakeholders' concerns to partition the CT conceptual domain in stakeholder-oriented sub-domains. The concepts found in the mental representation of the stakeholders' beliefs populate the universe of discourse and compose their vocabulary. In this way a concern-oriented ontology of the CT can be obtained. The method is illustrated for the subject selection criteria, a component of a CT project, but it can be used for any other CT component.

Keywords: clinical trial, selection criteria, ontology, concern

Introduction

A clinical trial (CT) is a "controlled study performed in human subjects and intended to discover, and/or evaluate, and/or verify safety, effectiveness, clinical and pharmaceutical effects, and adverse reactions of new drugs, devices, treatments, preventive measures, or other medical interventions in treating, preventing or diagnosing a specific disease or condition" [1]. The whole process involves high investments and several participating organizations, uses many resources, and is composed of numerous sub-processes, each of them carried out by agents playing specialized roles for the production of services and artefacts needed to the final evaluation of the research [2].

To reduce costs and enhance the research quality new approaches to the research project design and management as well as more IT support are needed. The current trend is to expand the IT support and automatize as many CT activities as possible. Nowadays, software products are available on the market for CRF (Case Report Form) gathering and processing, Clinical Trial Management, etc., however other CT

activities as protocol authoring remain largely manual because they are highly creative and intensely knowledge-based. Another bottleneck in the CT process is the lack of standardization and, as a consequence, of seamless interoperability between organizations, sub-processes, or agents at various levels of the research [3].

The aim of our research is to provide CT researchers with new methods and techniques able to supply the CT design, setup, and conduction with more effective IT support and enhance interoperability between the CT components [4]. We consider that a comprehensive ontology of clinical research can contribute to solve these issues achieving interoperability between software applications and information systems of CT organizations, extending automatization in new areas of the CT process, as the protocol authoring, and facilitating the CT standardization. Building such an ontology is a challenge for analysts and new methods to obtain it are welcomed.

The paper presents a concern-oriented approach to obtain the ontology of clinical trials by conceptualizing the CT domain from the point of view of stakeholders, that is persons having personal or professional interests in the development of the CT project and obtainment of accurate and trustworthy results. The concerns of all CT stakeholders partition the knowledge we have on the CT universe in smaller pieces which can be separately analyzed and described. Finally, this knowledge is mapped in a comprehensive ontology.

Ontologies

An ontology for clinical trials can help CT standardization enabling agreement on the meaning of CT vocabulary and facilitate interoperability and integration across applications. It can be represented in a specialized language enabling automatic querying and reasoning and in this way extending automatization of CTs. And the last but not the least, ontologies allow to integrate knowledge and data, so their use can help the development of more powerful and more interoperable information systems for CTs, especially based on service-oriented architectures. Such systems can reuse and share patient electronic records and integrate legacy systems with new ones in an enterprise integrated system.

In our research we use DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [5] with its extension D&S (Descriptions and Situations) [6,7] as a "foundational"[1] ontology where we "plug in" our CT ontology. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to particulars[2]. We adhered to the DOLCE +D&S ontology because:

- DOLCE captures the ontological categories belonging to natural languages and human commonsense;
- being a formal ontology [8], DOLCE and D&S precisely define its notions and relations used;

[1] A foundational ontology characterizes general terms (e.g. entity, event, process, spatial and temporal location) and basic relations (e.g. part-of, quality-of, participation, dependence).

[2] Particulars are entities that cannot have instances that are individuals, whereas universals represent ideas or abstract entities which can have instances.

- DOLCE and D&S contain all categories and relations we need for writing our CT ontology.

Concerns

Individual stakeholders in CTs can be classified depending on their role. A *role* is a unit of defined responsibility that may be assumed by one or more individuals (human being, team, or organization) in a collaborative environment aimed to deliver an artifact or service. The role is determinant for the concerns the stakeholder exhibits during the environment evolution. Our claim is that a concern is an intentional state of mind arising from preoccupation or interest in a *problem* which the stakeholder has identified in the real world.

Stakeholders and their concerns

There are many roles of stakeholders in a CT: sponsor, investigator, subject, data manager and so on. Sometimes it is useful to consider collective stakeholders as the writing committee formed by individual stakeholders, such as CT coordinator, statistician, etc.

In [9] we defined the (stakeholder's) concern as a problem-related care of one or more stakeholders involved in the construction or operation of an artefact in its natural environment. The care can derive from stakeholders' interests, desires or preoccupations for the environment's evolution.

The problem specification is the pair: a) *initial state* description (the current situation as the stakeholder perceives it) and b) *final state* description (the situation which matches expectations or goals of the stakeholder). These two elements are respectively considered as hypothesis and conclusion of the match-to-expectations tentative plan the stakeholder will have to fulfil. If grouped with the role of the stakeholder who manifests interest or preoccupation about the problem, this pair becomes the high level specification of a concern that the stakeholder will try to solve. We consider a problem's initial state containing all data, information, and knowledge necessary to obtain the final state of the problem. For example, design and management of the CT are based on the following two concerns which reflect the two main preoccupations of the CT managers: a) find out if the new drug or procedure is effective and guarantee the safety of the subjects enrolled in the CT and b) test the effectiveness of the new therapy and/or procedure.

C1	*Name*: Care not to worsen the general clinical state of the subjects during the clinical trial and follow-up	
	Problem	*Hypothesis*: Objectives were defined in order to start a clinical research on a new drug and/or procedure.

		Conclusion: The drug or procedure risks were minimized as much as possible and in the case of appearance of adverse events, their gravity was minimized, the safety of the subjects was preserved during the study and follow-up.
	Stakeholders: Sponsor, Scientific Coordinator	
C2	*Name*: Care to attain the objectives of the clinical trial	
	Problem	*Hypothesis*: Objectives were defined in order to start a clinical research on a new drug or procedure.
		Conclusion: The clinical trial objectives were attained.
	Stakeholders: Sponsor, Scientific Coordinator	

While a stakeholder acts for solving a concern, situations that trigger new concerns may occur. Between these concerns and the original one a causal relation exists. For instance, the concerns beginning with C4-C7 derive from the stakeholder's preoccupation to solve C3 (see bellow). But it is also possible that the natural evolution of the environment where the stakeholder operates to solve a concern generates new situations of concern without an apparent causal relation with another concern. The new concerns could be predictable or unexpected ones. They should be considered and solved by the stakeholder or delegated to other stakeholders responsible for solving them.

Towards an Ontology of the CT

In [9] a concern-oriented method for information system analysis is proposed. The method has five steps: 1) identification of stakeholders; 2) identification of concerns; 3) concern classification; 4) identification of relations between concerns; and 5) concern prioritization. The method also includes guidelines for designing an ontology of the domain under analysis. For this, seven steps should be added to the basic method: 6) identification of semantic rationales in concern specifications; 7) identification of the concepts used in the semantic rationales; 8) ontological analysis of the intension (connotation) of the concepts; 9) choosing a foundational ontology to be extended by our ontology; 10) classification of the concepts conforming the foundational ontology; 11) identification of inter-concept (conceptual) relations in concepts' semantic rationales; and 12) definition of the ontology using a formal logic language. Such an ontology is structured according to stakeholders' responsibilities and concerns through the identification of shared and specific vocabularies.

In the followings we present this method for the sample case of an ontology of the selection criteria of CT subjects. In this case the concerns we consider are mainly related to the writing committee stakeholder.

Concerns identification

In order to identify the concerns and their dependencies, our method firstly recommends to analyse the stakeholders' preoccupations, interests and beliefs, and identify how they generate concerns, in other words how the stakeholders reason. For this, we resort to contemporary philosophy that gives us many theories of mind.

The most known and accepted theory is the Functionalism which models the states of mind (beliefs, concerns, desires, being in pain, etc.) by considering solely their functional role: transformers of sensory inputs in behavioral outputs, in causal relations with other states of mind [10]. States of mind are closely related with knowledge, that is information evaluated, processed, and structured by the human mind in conclusions or explanations so that it can be used according to current needs. They are based on mental representations that contain knowledge structures.

In our method, the beliefs take a special place because they are closely related to concerns. They represent convictions a stakeholder holds for true in a given situation, independently of the nature of the source of the conviction.

One of the first tasks in the CT process is to define the person characteristics needed to select the study population that is a homogeneous and reliable number of subjects identified on the basis of a statistical analysis on a specific pathology. A complete, consistent, and non-redundant set of characteristics is the primary condition to obtain trustworthy results. These characteristics are used to describe inclusion criteria, that is criteria that prospective subjects must meet to be eligible to participate to the study, and exclusion criteria, that is, criteria that may exclude a potential subject from participation in the study, even if this subject satisfies the inclusion criteria. One compulsory exclusion criterion is the not-signing of the informed consent. The subject selection criteria are an aggregation of inclusion and exclusion criteria.

In this paper we focus our analysis on the subject selection criteria and their main stakeholder, the writing committee, which has the following concern regarding the identification of the subject selection criteria:

C3	*Name*: Care to obtain the characteristics of the study population	
	Problem	*Hypothesis*: Following information is available: 1. Trial objectives, trial design, statistical considerations 2. Previous CT projects and their scientific results *Conclusion*: Characteristics of the CT subjects are identified and specified as inclusion/exclusion criteria.
	Stakeholders: Writing committee	

Other concerns, C4-C7 are derived representing subtasks of the comprehensive task represented by C3. For simplicity of presentation in the followings the concerns are specified only with their code and name. For deriving the new concerns the members of the writing committee make use of tacit knowledge and their work practice (for example, design methodologies, design guidelines, operative manuals, etc).

In the following table four samples of knowledge are specified through an identification code and a description in natural language of their corresponding mental representations.

Code	Mental representation description in natural language
K1	The subjects who have the disease in the stage considered by the study are eligible.
K2	The subjects who have other medical condition that could bias the study compliance and follow-up are not eligible.
K3	The therapeutic procedure administration will affect the clinical manifestation of the disease.
K4	Results of previous phases of the CT or experimental laboratory results are available and a list of ascertained side effects and/or adverse events is known.

According to K1, the set of possible CT participating subjects considers those who have the disease in the stage under study. This is why, in order to solve the concern C3, a first care is to find how the clinical manifestation of the disease is related to the objectives attainment and the selection criteria identification, that is the concern C4:

Code	Concern name
C4	Care to find how the clinical manifestation of the disease is related to the objectives attainment and the selection criteria identification

From the CT general objective as well as from the concerns C1-C4 and from the knowledge pieces K1-K4 originate the following concerns:

Code	Concern name
C5	Care to determine how the components of the study therapeutic procedure and its administration might affect the general state of the subjects and/or influence the attainment of the objectives
C6	Care to determine the way in which already known adverse events of the study therapeutic procedure administration might affect the general clinical state of the subjects and/or influence the attainment of the objectives

Other concerns derive from the analysis of particular views of the selection criteria, for example the ethic aspects originated in C1. For instance the next concern, C7, comes from the responsibility of the writing committee to specify selection criteria of quality:

Code	Concern name
C7	Care to specify unambiguous, consistent, and not redundant selection criteria as a sum of inclusion and exclusion criteria

Beliefs Related by Concerns

Our method firstly proposes an analysis of the stakeholders' beliefs (but also their preoccupations or interests) for identifying the causal relations between states of

mind. Mental representations of the beliefs are expressed in natural language as we did for pieces of explicit knowledge. Table 1 collects the results of the analysis of the concerns C1-C4. In the table "Bx" is used for coding mental representation which corresponds to a belief where B stands for a belief and x is a number.

State of mind code	Mental representation description in natural language
B1	There is the risk that the general clinical state of the subjects may become worse during the clinical trial and follow-up
B2	The safety of the subjects is preserved during the clinical trial and follow-up
B3	The objectives of the clinical trial are fulfilled
B4	The study population is formed by subjects.
B5	The study population has characteristics
B6	The characteristics of the study population are defined by the inclusion and exclusion criteria.
B7	The inclusion and exclusion criteria form the selection criteria.
B8	The clinical manifestation of the disease has relation with the attainment of the objectives and the selection criteria identification
B9	The disease in its stage has a diagnostic as it resulted from tests
B10	The existence of the disease and its gravity affect the life expectancy of a person
B11	The life expectancy is a characteristic of study population
B12	The study population has other basic characteristics: gender, race and performance status
B13	The existence of the disease and its gravity affect the performance status of a person
B14	The clinical manifestation of the disease might have some particularities that are important to identify the selection criteria.
B15	Any disease has a pathogenesis.
B16	The pathogenesis of the disease might affect the general clinical state of the subjects and the inference of the selection criteria.

Table 1 Beliefs which originate concerns C1-C4

Dependency in the Concern Space

We represent the causal relations between intentional states of mind (beliefs, preoccupations, etc.) as dependencies. They may induce dependence relations between concepts.

There are two kinds of dependence between states of mind: direct and bound.

Direct dependence. It indicates a causal relation describing that the origin of one component of the relation resides in the other component. The relation may relate: a) a concern and its relative belief, preoccupation, interest; and b) two beliefs, preoccupations, interests, and knowledge.

The first case can be exemplified with the concern C1 which is derived from a preoccupation P1: “the general clinical state of the subject does not be worsen during the clinical trial and follow-up”. In its turn P1 causally depends on the belief B1.

The second case comes out when there are beliefs which cause preoccupations or interests on the same problem considered from two different points of view. For example, the knowledge K4 says that adverse events of the study therapeutic procedure administration have been ascertained by previous tests and/or research studies cause the preoccupation to minimize them during the study. This preoccupation expresses the preoccupation P1, the same we used above, but from the point of view of the adverse events occurrence risk. The risk could be mitigated if, for instance, we require a normal function of major organs (e.g. cardiac function at least Grade 2). These requirements contribute to building of the inclusion criteria for subjects’ enrolment.

To explain the direct dependence relation between beliefs, let consider the beliefs that appear during the concern C4 and their relations (Fig. 1). Because C4 depends (historically) on the belief B8, we can suppose that the reasoning process begins with the state B8. This state gradually originates mental representations of the contained concepts: clinical manifestation of a disease, CT objective, selection criterion, etc. In the context of these representations, the disease clinical manifestation concept is the most important. It is related to the representation of the disease concept, due to the link of concept- property type between disease and clinical manifestation. This link can also be inferred from the semantic net of the concepts of the belief B8.

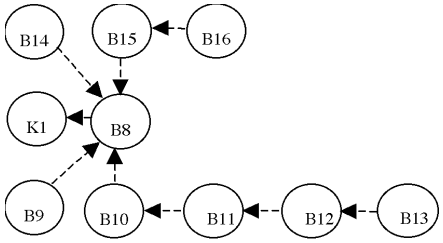


Fig. 1 Direct dependence relations between the beliefs related to C4

Using the same reasoning process, we can infer that the mental representation of the disease concept is bound to representations of other concepts (generated on knowledge), like diagnostic, life expectancy of a person, and pathogenesis.

The link between disease and diagnostic is expressed in the belief B9 which expresses the fact that at enrollment time the disease of the potential subject is diagnosed and a diagnostic was delivered after some tests have had been carried out.

If we re-do the reasoning process, we can conclude that the belief B8 (the starting concept) caused the state B9, otherwise the B9 state would not exist if the B8 state did not exist. This is in fact the dependence relation between B9 and B8.

Similarly, if we apply the last two steps in the case disease-life expectancy of a person and disease-pathogenesis, we obtain dependency relations between B10 and B8, respectively B15 and B8. These relations are showed in Fig. 1.

Bound dependence. This relation is between two states of mind (interests, needs or desires, and beliefs), and expresses the fact that one state of mind is based on the

other. For instance, the concern C2 causally depends on an interest I1 "the objectives of the clinical trial to be fulfilled". The mental representation expressed by this sentence is the same as the mental representation of the sentence "the objectives of the clinical trial are fulfilled" which is related by the belief B3. At the ending of the clinical trial, this sentence is true or not, depending on whether the CT was successfully closed. In this case we use the bound dependence to express the fact that the mental representation of an interest is the same or has the same semantic as the mental representation of a belief.

Sample of CT Ontology Definition

In Table 2 concepts derived from the concerns C1-C3 and regarding the selection criteria definition are defined. Some of them are CT specific (selection criteria, inclusion criteria, exclusion criteria), while others (person, variable, value of variable, maximal domain) belong to some related ontologies.

Concept	**Formal semantic**	**Description**
person	Person(p)$=_{df}$ APO(p)$\wedge\exists$c (DNA(c)$\wedge$ OSD(p,c))	Agentive physical object that specifically and one-side depends on DNA.
selection criteria	SelectionCriteria(c)$=_{df}$ Description(c)$\wedge$ $\exists$w,t(WritingCommittee(w)$\wedge$ Conceives(w,c,t))$\wedge\exists$i1,i2,e,t (InclusionCriterion(i1)$\wedge$ InclusionCriterion(i2)$\wedge$ ExclusionCriterion(e)$\wedge$ TimeInterval(t)$\wedge$ K(i1,c,t)$\wedge$ K(i2,c,t) $\wedge$K(e,c,t))	Description conceived by the writing committee and constituted by the inclusion and exclusion criteria. The description has two parts: inclusion and exclusion criteria. Moreover, the persons have to have signed and dated an informed consent, K1 and K2.
variable	(A) Variable(v) $\rightarrow$ Description(v) (A) Variable(v)$\wedge\forall$p1,s1 (Particular(p1)$\wedge$Situation(s1)$\wedge$SAT(s1,v)) $\rightarrow\exists$p2,s2, t (Particular(p2)$\wedge$ TimeInterval(t)$\wedge$(P(p1,p2)$\vee$PC(p2,p1,t)$\vee$ qt(p1,p2))$\wedge$Situation(s2)$\wedge$P(s1,s2)$\wedge$ SettingFor(s2,p2)$\rightarrow\exists$d (Description(d))$\wedge$ SAT(s2,d)$\wedge$PP(d,v))	Generic description of various situations of a particular (namely endurant, perdurant or quality) used in the description of another particular. Such variable exists due to a kind of relation between the two particulars. The relation can be: parthood, participation, or quality.
value of variable	(D) ValueOf(s, v)$=_{df}$ Situation(s)$\wedge$Variable(v)$\wedge$SAT(s,v) (A) Variable(v) $\rightarrow\exists$s1,s2 (Situation(s1)$\wedge$ Situation(s2)$\wedge$ValueOf(s1,v)$\wedge$ ValueOf(s2, v)$\wedge$ $\forall$x1,x2(Particular(x1)$\wedge$SettingFor(s1,x1) $\wedge$Particular(x2)$\wedge$SettingFor(s2,x2)$\rightarrow$ x1$\neq$x2))	A situation that satisfies a variable description. A variable can take different values.
maximal domain	MaximalDomain(v)$=_{df}$ Variable(v)$\wedge$	Domain of all values of a variable.

Concept	Formal semantic	Description
	∀s (Situation(s)∧ SAT(s,v) → valueOf(s,v))	
characteristic of persons	(A) CharacteristicOfPerson(p,v) → Person(p)∧Variable(v) (A) CharacteristicOfPerson(p,v)→ ∃v1(Variable(v1)∧ CharacteristicOfPerson(p,v1)∧PP(v1,v)) (A) ¬∀p, v(Person(p)∧Variable(v)∧ CharacteristicOfPerson(p,v)→ ∃v1(Variable(v1)∧ CharacteristicOfPerson(p,v1)∧PP(v1,v)))	Variable that applies on a person. In addition, the characteristics of a person can be parts of other characteristics.
assessment domain	(A) AssessmentDomain(v) → ∃d(Variable(d)∧ MaximalDomain(d)∧ PP(v,d)) (A) AV(s,v) → ValueOf(s,v) (A) ∃s,v(Situation(s)∧Variable(v)∧ ValueOf(s,v) → ¬AV(s,v)	Sub-domain of the maximal domain which contains the values that are considered in the study.
inclusion criteria	(D) InclusionCriteria(d, p)=$_{df}$ Description(d)∧Person(p)∧ ∃v(Variable(v)∧ CharacteristicOfPerson(p,v)∧PP(v,d)) (A) ∀d,p(Description(d)∧Person(p)∧ InclusionCriteria(d, p)→ ∃s(Situation(s)∧AV(s,d))	Description conceived by the writing committee of a non-empty group of assessment domain for a selection list of characteristics of persons. It is used to select (eligible) persons for the CT.
exclusion criteria	(D) ExclusionCriteria(d, p)=$_{df}$ Description(c)∧ Person(p)∧ ∃v(Variable(v)∧ CharacteristicOfPerson(p,v)∧PP(v,d)) (A) ∀d,p(Description(d)∧Person(p)∧ ExclusionCriteria(d, p)→ ∃s(Situation(s)∧¬AV(s,d))	Description conceived by the writing committee of a non-empty group of assessment domain for a selection list of characteristics of persons. It is used to eliminate persons from the eligible ones.

Table 2 Excerpt from the ontology of the subject selection criteria

From Table 1 and Table 2 the concepts belonging to the concerns C1-C4 are gathered and their taxonomy is represented according to DOLCE+D&S in Fig. 2.

Conclusions

In this paper, a concern-oriented method for designing an ontology is presented. The method is based on stakeholders' concerns to partition the CT conceptual domain in stakeholder-oriented sub-domains. The stakeholders' concerns come from their interests, desires, and preoccupations and depend on their beliefs and knowledge.

Mental representations of stakeholders related to each concern are identified as clusters of concepts related to each other. We consider such a cluster as the semantic rationale of the associated concern. The concepts found in the semantic rationales populate the universe of discourse specific to each stakeholder and compose the stakeholder's vocabulary. Some concepts are shared with other stakeholders while others are specific to one stakeholder; some concepts are domain-specific while others are medical or general concepts.

The method is illustrated to derive the ontology of a component of the clinical trial protocol, the subject selection criteria but it can be used for any other component of the CT. Our current research is focused on the conceptualization of the CT protocol, the planning document where all core concepts of the clinical trial are found, in order to encourage standardization and improve interoperability in clinical research.

References

1. NCI Thesaurus: http://nciterms.nci.nih.gov
2. Collada, A., Fazi, P., Luzi, D., Ricci, F. L., Serbanati, L. D., Vignetti, M.: Toward a Model of Clinical Trials. Biological and Medical Data Analysis. 5th International Symposium, ISBMDA, Barcellona, Spain, Lecture Notes in Computer Science, Springer-Verlag, Berlin (2004)
3. Fazi, P., Collada, A. L., Luzi, D, Ricci, F. L., Serbanati, L.D., Vignetti, M.: A Proposed Clinical Trial Model: Analysing the CT Process. *Applied Clinical Trials*, Vol. 15, 1, (2006)
4. Luzi, D., Ricci, F.L., Serbanati, L.D.: E-Clinical trials supported by a service-oriented architecture. To be published in MedNet 2006: The 11th Word Congress on Internet in Medicine, Toronto (2006)
5. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A: WonderWeb Deliverable D18. Ontology Library. IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web (2003)
6. Gangemi, A., Catenacci, C., Lehmann, J., Borgo, S.: Task Taxonomies for Knowledge Content, EU 6FP METOKIS Project D07 (2005)
7. Gangemi, A., Mika, P.: Understanding the Semantic Web through Descriptions and Situations", International Conference ODBASE03, Italy, Springer (2003)
8. Guarino, N.: Formal Ontology and Information Systems, Proceedings of Formal Ontology in Information Systems (FOIS 1998) IOS Press, Trento, Italy (1998)
9. Bogdan, C., Serbanati, L. D.: Toward a Concern-Oriented Analysis Method for Enterprise Information Systems, IEEE International Multi-Conference on Computing in the Global Information Technology (ICCGI 2006), Bucharest, Romania (2006)
10. Block, N., ed.: Readings in Philosophy of Psychology, 2 vols. Vol. 1, Cambridge, Harvard (1980)

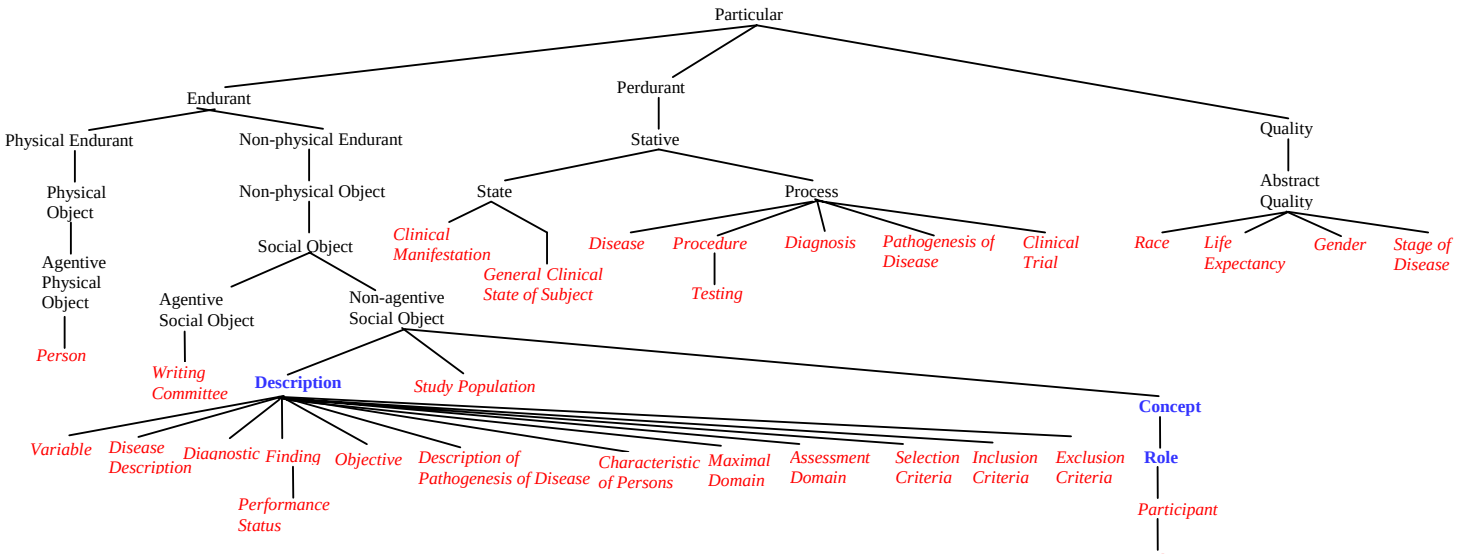


Fig. 2 Classification according to DOLCE + D&S of the concepts that belong to the concerns C1-C4