

# Web Ontology Language Requirements w.r.t Expressiveness of Taxonomy and Axioms in Medicine

Christine Golbreich<sup>1</sup>, Olivier Dameron<sup>2</sup>, Bernard Gibaud<sup>2</sup>, and Anita Burgun<sup>1</sup>

<sup>1</sup> Laboratoire d'Informatique Médicale,  
Faculté de Médecine, Av du Pr. Léon Bernard, 35043 Rennes France  
Christine.Golbreich@uhb.fr,  
Anita.Burgun@univ-rennes1.fr

<sup>2</sup> Laboratoire IDM, UPRES-EA 3192  
Faculté de Médecine, Av. du Pr. Léon Bernard, 35043 Rennes Cedex France,  
{Olivier.Dameron, Bernard.Gibaud}@chu-rennes.fr

**Abstract.** An important issue is to know whether Web ontology languages, meet the expected requirements of expressiveness and reasoning. This paper aims at contributing to this question in evaluating and comparing several languages. After describing the needs of a Semantic Web in medicine, it analyses Protégé and DAML+OIL primitives on a concrete medical ontology, the brain cortex anatomy ontology. It draws conclusions about the requirements that a Web ontology language should meet for the representation of medical taxonomy and axioms. The expressiveness of DAML+OIL or OWL DL seems suited to describe the complex taxonomic knowledge. But rules are required for representing the deductive knowledge (dependencies between relations) and to support several tasks (ontology construction, maintenance, verification, query of heterogeneous distributed information sources). Finally, the paper evaluates the features of the next standard OWL and of an hybrid language CARIN-ALN with respect to these requirements.

## 1 Introduction

A major challenge for the Web is to evolve towards a « Semantic Web », in which information may have explicit semantics, enabling human and machines to make a better use of information, and better integrate available data. The semantic markup of data is a means to reach this goal. Ontologies play a central role in the Semantic Web, since they define the vocabulary for such semantic markup. Thus an important issue is to know whether Web Ontology Languages, meet the requirements of expressiveness and reasoning that may be expected from the Web communities. This paper aims at contributing to this question in evaluating two ontology languages, Protégé [15] and DAML+OIL[5] by using them to represent a concrete ontology in the medical domain, the brain cortex anatomy ontology. This experiment pointed out some lacking of expressiveness that a Web Ontology Language shall overcome and enabled to state important features to provide w.r.t expressiveness of the taxonomy and axioms, for meeting the biomedical community usual requirements.

Section 2 gives a brief overview of the major Web uses expected from the biomedical community. Section 3 presents motivation for a Web brain cortex ontology. Section 3 presents the main features of Protégé and DAML+OIL, the precursor of the future W3C standard OWL [18]. Section 5 analyses Protégé and DAML+OIL expressiveness on the brain cortex anatomy ontology and draws conclusions about the requirements that a Web ontology language should meet for the representation of medical taxonomy and axioms. Section 6 compares formal languages for representing medical ontologies and discusses requirements of the biomedical community for a Semantic Web Ontology language.

## 2 Biomedical Community Web Uses

Semantic Web shall improve existing Web-based applications and enable new uses of the Web for the biomedical community. Searching and getting easily information on the web, and using it for decision making are the major needs regarding the web in medicine.

Search on the web concerns all actors of the medical field: medical doctors, researchers, patients, as well as manufacturers e.g. from the pharmaceutical or medical device industry. Current search engines are mostly based on the use of keywords, and therefore may return non relevant information, due to homonymy, synonymy, hyponymy-hypernymy etc. The solution retained in documents repositories is to index and search documents using « descriptors » from a thesaurus, rather than keywords. This is the approach actually used with thesauri such as Digital Anatomist [22], or MEDLINE<sup>1</sup>, based on the MeSH thesaurus. Existing thesauri and languages like the Unified Medical Language System UMLS<sup>TM</sup> [14] or the Medical Subject Headings, MeSH<sup>2</sup> provide a very significant basis that cannot be ignored. However, they have limitations. The lack of precision of the concepts definitions may lead to non-shared meanings, which may jeopardize reuse and interoperability. Some highly specific concepts may be absent (cf. § 3). The rapid evolution of medical knowledge and the very dynamic nature of web information require frequent updates, enforcing a strict version management, as well as detection of inconsistencies that might result of updates or modifications.

Medicine has now evolved to a more scientific discipline, highly specialized, and therefore exercised collectively. Thus, the sharing of medical data of increasing volume and complexity has become critical to guarantee seamless healthcare delivery. The setting up of medical data repositories, centralized or federated on the web, and articulated around common ontologies, is of critical importance for modern research. This is particularly crucial in multi-disciplinary domains of research – such as neurosciences [26], [3] – that require the sharing of both knowledge and data. The feasibility of this sharing assumes the proper alignment of several domain ontologies (e.g. clinical neurology, psychiatry, neuroimaging, anatomy, genomics, neural models, neurochemistry). The case of imaging is also important, since signals and images need to be described, in order to make their content and context e.g. their

---

<sup>1</sup> <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

<sup>2</sup> <http://www.nlm.nih.gov/mesh/>

acquisition protocol explicit. Biological molecular and Genomics databanks are another obvious illustration.

Moreover, the problem is not limited to information sharing. What is at stake is also to be able to *use* this information for decision making. This also includes many aspects: for instance access to clinical guidelines and protocols (patient management, prescription) and reference to optimal clinical management (evidence-based medicine), which are fields where the Semantic Web should bring - at a reasonable price - meaningful contributions [23]. One should also mention decision support brought by knowledge-based systems, which is much harder to implement, because it requires medical data in machine-understandable form, as well as structured knowledge, in order to adequately interpret and process this data.

The Semantic Web languages should contribute to better address those needs. A formal language with a clear semantics for representing ontologies (§ 4.2), should allow a formal definition of concepts, facilitating precise and shared concepts meaning, automatic consistency checking, automatic concepts classification. Its extension by a rule formalism (§ 5.2), should allow the definition of inferences leading to a more powerful search, as well as the query of heterogeneous and distributed information sources.

### 3 Brain Cortex Anatomy Ontology

The use of anatomy as a ground to model many fields of medicine [22] is a first motivation for a Web ontology of anatomy, that could be shared as a common reference in these fields [3]. Modeling anatomy has been pursued in several works such as Galen [21] or Digital Anatomist [22]. However, they are much focused on the general level of organs and do not provide detailed descriptions of specific organ structures. They are missing a fine-grained description of the brain cortex. Moreover, a general ontology of brain cortex anatomy should be reusable in various applications: teaching, decision support for clinical practice, and neuroimaging data sharing among collaborating research centers, for instance in order to improve the statistical significance of research findings through the use of larger populations of subjects. Therefore, another reason to achieve a Web ontology of anatomy is to provide a model of anatomy independent from application, that could be usable by humans as well as for different software. For all these reasons, it is important to develop a Web ontology of brain cortex anatomy.

Modeling brain cortex anatomy requires to explicitly define the meaning and properties of the domain concepts and their relations. The definitions of the anatomical concepts in the following examples are based on anatomy atlases such as [17], and terminology sources such as NeuroNames [2]. For instance, a «brain hemisphere» is informally defined as an anatomical part of the cortex which is lateralized (i.e. located either on the right or on the left side), which includes five anatomical subdivisions called lobes (frontal, temporal, parietal, occipital and limbic lobes) and occupies a specific region of space. A medical ontology like the brain cortex ontology involves many different relations: specialization (e.g. frontal lobe is a lobe), part-of (e.g. opercular pars of inferior frontal gyrus is a part of inferior frontal gyrus), topological, causal relations etc. Representing the brain cortex relations arise some difficulty because they are not independent. For instance, dependencies between

composition and topological relationships have been studied in [27], while propagation of relations along transitive role, for instance location of a disease is inherited across patrimony: “has-location propagates -via part-of” is analyzed in [20]. In order to insure consistency, a stratified representation of brain cortex anatomy with a spatial stratum representing space regions occupied by anatomical entities has been advocated in [4]. In that way, the relations between anatomical concepts can be defined from topological relationships associating the regions they take up. With these definitions, the symbolic model of anatomy can be related to anatomical imaging data, which in turn can be used as localization support for functional activities or pathological elements. Furthermore, dependencies between anatomical relations can be derived from properties of topological relations, and could be automatically inferred. But this requires a formalism that allows the explicit representation of terminological knowledge (concepts, relations, and taxonomical organization) with classification services and also of deductive knowledge (axioms) with capability of handling inferences.

## 4 Representation

Two representation languages have been used to represent the ontology, a frame-based language supported by the Protégé-2000 editor, and DAML+OIL based on the description logics *SHIQ* supported by the OILED editor.

### 4.1 Ontology in Protégé

Knowledge representation in Protégé is based on frames. Protégé-2000 is a graphical and easy-to-use ontology-editing tool developed at Stanford University. The class inheritance hierarchy is visualized as a tree (Fig.1) multiple inheritance is allowed. Users define and organize classes, subsumption relationships, properties and property values. Metaclasses and hierarchy of metaclasses can be defined. A UMLST<sup>TM</sup> widget has been developed [13]. It allows users who are developing and populating their

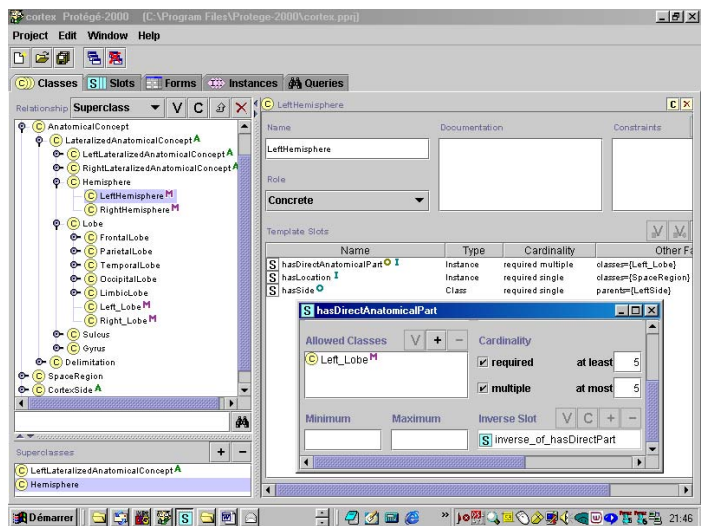


Fig. 1. Class hierarchy and LeftHemisphere definition

knowledge bases in Protégé to search and import UMLS™ elements directly into Protégé-2000. Other on-line resources can be used in a similar manner for knowledge acquisition in Protégé, e.g., WordNet [14].

For example, a «brain hemisphere» is defined as an anatomical part of the cortex which is lateralized (i.e. located either on the right or on the left side), includes five anatomical subdivisions called lobes (frontal, temporal, parietal, occipital and limbic lobes) and occupies a specific region of space. «Left hemispheres» are represented in Protégé by the class `LeftHemisphere`, subclass of `Hemisphere` and of `LeftLateralizedAnatomicConcept` whose slots are inherited but some of them overloaded: slots `hasSide` restriction: `LeftSide`, `hasDirectAnatomicalPart` restriction: `LeftLobe`, `facets` at least, at most with value 5 (Fig. 1), etc. This representation only expresses that a hemisphere has 5 lobes, each types confused. It would be difficult with frames (but possible) to represent that a hemisphere has exactly one lobe of each type, frontal, temporal, etc. (Ex14 gives the DAML+OIL representation).

## 4.2 Ontology in DAML+OIL

DAML+OIL [5] is a semantic markup language for Web resources designed by the W3C OntoGroup (Ontology Working Group) in order to go beyond the simple presentation of information on the Web, by offering a solution that supports interoperability, understanding and reasoning with these information. DAML+OIL, comes from DARPA results about DAML (American Agent Markup Language) [11] and from OIL [8]. DAML+OIL borrows its intuitive modeling primitives to frames, its syntax to the Web languages XML and RDF, its formal semantics and reasoning to description logics (DLs). From a formal point of view, DAML+OIL can be seen to be equivalent to the expressive description logic *SHIQ* with the addition of the `oneOf` constructor and datatypes (Table 1). It also supports the definition of a set of axioms (Table 2). DAML+OIL can make use of the Fast Classification of Terminology system FaCT (<http://www.cs.man.ac.uk/~horrocks/FaCT/>), which enables to reason on ontologies (consistency checking, classification). DAML+OIL, is the precursor of the future W3C Web Ontology Language OWL [18], intended “to facilitate machine readability of web content than that supported by XML, RDF, and RDF Schema, by providing additional primitives along with a formal semantics”. OWL provides three increasingly expressive sublanguages: OWL Lite, OWL DL, and OWL Full. OWL Lite is less expressive, thus according to its designer, “it should be simpler to provide tool support for OWL Lite than for the two others, and easier to provide a quick migration path for thesauri and other taxonomies”. OWL-DL offers completeness and decidability. OWL Full with its maximal expressivity and syntax freedom of RDF, offers no computational guarantees.

OILED is a graphical ontology-editing tool developed by the University of Manchester. It allows users to design ontologies in DAML+OIL [5]. The OILED editor is based on the DAML+OIL model. While OILED uses graphical user-interface metaphors common to frame-based systems, it offers a more expressive ontology language based on Description Logics, including reasoning services. Users define classes, subsumption relations, and properties with type restriction (Fig. 2). Opposed to frames editors, complex descriptions can be used as slot value, without having to

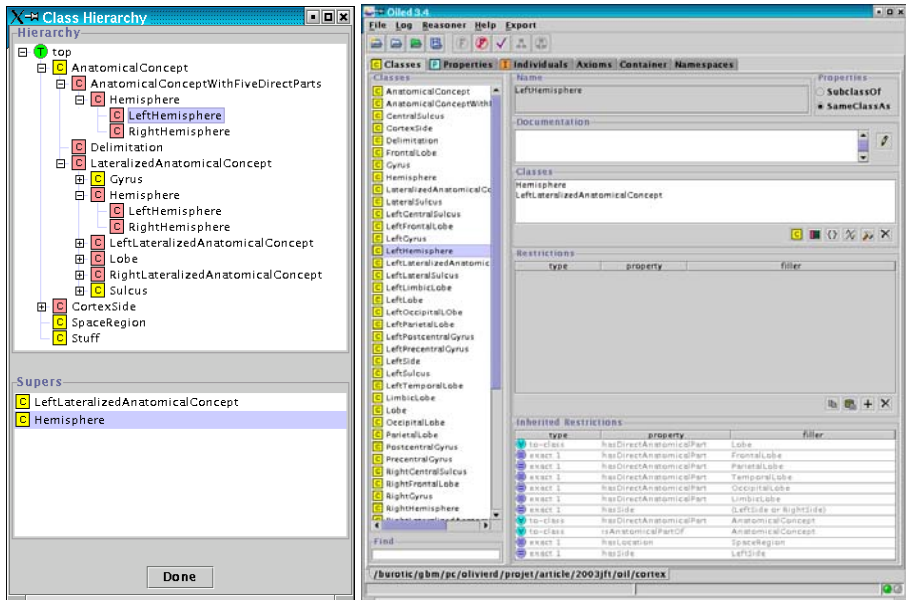


Fig. 2. Left: Post-classification hierarchy - Right : LeftHemisphere definition with OILED

be named. Axioms allow for representing additional knowledge, e.g. the fact that two classes are disjoint.

The next examples illustrate the rigorous formalization of concepts and taxonomy in DAML+OIL, and their automatic classification supported by FaCT.

- Ex1.** /An anatomical concept is composed of direct parts, which are anatomical concepts, occupies exactly one region of space  
 $AnatomicalConcept := (\forall hasDirectAnatomicalPart AnatomicalConcept) \wedge (= 1 hasLocation SpaceRegion)$
- Ex2.** /A lateralized concept is located either on the right side, or on the left side of the cortex, one can distinguish right-sided lateralized concepts and left-sided lateralized concepts/  
 $LateralizedAnatomicalConcept := AnatomicalConcept \wedge (= 1 hasSide LeftSide \vee RightSide)$   
 $LeftLateralizedAnatomicalConcept := LateralizedAnatomicalConcept \wedge (\forall hasSide LeftSide) \text{ resp. } RightLateralizedAnatomicalConcept$
- Ex3.** /An hemisphere is a lateralized concept whose direct parts are lobes, each part being of a distinct type (i.e. frontal lobe, parietal lobe, occipital lobe, limbic lobe, temporal lobe/  
 $Hemisphere := LateralizedAnatomicalConcept \wedge (\forall hasDirectAnatomicalPart Lobe) \wedge (= 1 hasDirectAnatomicalPart FrontalLobe) \wedge (= 1 hasDirectAnatomicalPart ParietalLobe) \wedge (= 1 hasDirectAnatomicalPart OccipitalLobe) \wedge (= 1 hasDirectAnatomicalPart LimbicLobe) \wedge (= 1 hasDirectAnatomicalPart TemporalLobe)$
- Ex4.** /A left (resp. right) hemisphere (resp. lobe etc.) is an hemisphere (resp. lobe etc.) located on the left (resp. right) side Fig.4 /  
 $LeftHemisphere := LeftLateralizedAnatomicalConcept \wedge Hemisphere$

The `LeftHemisphere` concept is defined as a `Hemisphere` as well as a `LeftLateralizedAnatomicalConcept`, together with a number of restrictions on its direct parts (Ex3). Consequently, it has exactly 5 direct parts, which are: `LeftFrontalLobe` `LeftParietalLobe` `LeftOccipitalLobe` `LeftLimbicLobe` `LeftTemporalLobe`. Thus, the FaCT classifier automatically classifies it as subsumed by `FiveDirectPartAnatomicalConcept` as shown on the Post-classification hierarchy (Fig. 2), whereas it was firstly defined as only subsumed by `LeftLateralizedAnatomicalConcept` and `Hemisphere`. This example exhibits how a formal language like DAML+OIL allows a finer-grained description than Protégé and the benefits of automatic classification.

## 5 Requirements and Lacking of Expressiveness

The following examples from the brain cortex anatomy ontology present the main features that are covered by Protégé or DAML+OIL (§ 5.1), and the requirements of expressiveness that they do not meet (§ 5.2). Thus this study enables to draw first conclusions about the expressiveness expected from a Web language for medical ontologies. For each example, the number refers to the Protégé-2000 and DAML+OIL constructor or axiom e.g. subsumption, equivalence that have been used, denoting a specific requirement of expressiveness (§ 6 column 2 – 3 Table 1 Table 2).

### 5.1 Maximum Expressiveness Required

The following examples 1 to 14 exhibit that OWL Lite is not sufficient, and that the enhanced expressiveness of OWL DL is at least required for representing the brain cortex ontology, since for instance disjunction, negation, disjoint union are needed

- Ex5.** *disjointWith* is needed to represent disjunction (#9)  
 / Hemisphere, Lobe, Gyrus et Sulcus are disjoint classes /  
`disjointWith (Hemisphere Lobe Gyrus Sulcus)`
- Ex6.** *disjunction* is required (#2)  
 / Lateralized anatomical concept are either right or left /  
`LateralizedAnatomicalConcept := AnatomicalConcept`  
`^ (= 1 hasSide LeftSide v RightSide)`
- Ex7.** *negation* is needed (#5)  
 / Class of the anatomical concepts that are not lateralized /  
`NonLateralizedAnatomicalConcept := AnatomicalConcept`  
`^ ¬ LateralizedAnatomicalConcept`
- Ex8.** *disjointUnionOf* is a primitive needed to represent a partition of A into a list of concepts (#10)  
 / A side of the brain cortex is either right or left but not both /  
`disjointUnionOf (CortexSide LeftSide RightSide)`  
 / A lobe is one of the following types : frontal, parietal, temporal, occipital, limbic lobe /  
`disjointUnionOf (Lobe FrontalLobe ParietalLobe TemporalLobe`  
`OccipitalLobe LimbicLobe)`  
 / An hemisphere is either a right hemisphere or a left hemisphere /  
`disjointUnionOf (Hemisphere LeftHemisphere RightHemisphere)`
- Ex9.** *≡ (equivalent)* is needed to represent classes equivalence (#8)  
 / The left lobe concept is equivalent to left lateralized anatomical concept and lobe /  
`LeftLobe ≡ LeftLateralizedAnatomicalConcept ^ Lobe`

- Ex10.**  $\subseteq$  subsumption is needed to represent class or relation specialization hierarchies (#11)  
 /The relation hasAnatomicalPart is a specialization of hasDirectAnatomicalPart/  
 hasDirectAnatomicalPart  $\subseteq$  hasAnatomicalPart
- Ex11.** transitive is needed for representing transitive relations (#14)  
 /has-part is transitive (hasDirectPart no)/  
 Slot-def has-part TransitiveProperty  
 Representing such property features (reflexivity, symmetry, transitivity) is required. For example, transitivity enables to elicit the distinction between hasDirectAnatomicalPart and hasAnatomicalPart. The latter corresponds to the transitive closure of hasDirectAnatomicalPart, e.g. direct anatomical parts of hemispheres are lobes, thus anatomical parts of hemispheres are lobes and gyri. DAML+OIL provides such a possibility while Protégé does not.
- Ex12.** **inverse relation** is needed (#13)  
 /inverse of hasLocation/  
 isLocatedIn inverseOf hasLocation
- Ex13.** **equivalence of relations** must be represented (#12)  
 /concept A is anatomical part of a concept B if and only if the space occupied by A is a subspace of that occupied by B/  
 isAnatomicalPartOf  $\equiv$  (isLocatedIn  $\circ$  isSubAreaOf  $\circ$  hasLocation)  
 From this definition, constraints on body spaces can be inferred for two anatomical concepts A and B linked by isAnatomicalPartOf, and inversely A isAnatomicalPartOf B can be inferred from their respective regions. Moreover, equivalence between relations is crucial to merge several Web ontologies.
- Ex14.** **cardinality** and non exclusive constraints on relations have to be represented ( $n^{\circ 6}$ )  
 /An hemisphere is a lateralized concept whose direct parts are lobes and which has exactly one lobe of each type/  
 Hemisphere := LateralizedAnatomicalConcept  
 $\wedge$  ( $\forall$  hasDirectAnatomicalPart Lobe)  
 $\wedge$  (= 1 hasDirectAnatomicalPart FrontalLobe)  
 $\wedge$  (= 1 hasDirectAnatomicalPart ParietalLobe)  
 $\wedge$  (= 1 hasDirectAnatomicalPart OccipitalLobe)  
 $\wedge$  (= 1 hasDirectAnatomicalPart LimbicLobe)  
 $\wedge$  (= 1 hasDirectAnatomicalPart TemporalLobe)  
 DAML+OILs allow for representing such constraints whereas Protégé does not.

## 5.2 Needs of a Rule Layer

OWL DL is not yet sufficient. The next examples show needs that are not covered by Protégé and DAML+OIL, and exhibit that the expressiveness of rules is required. Indeed, the brain cortex ontology involves a lot of deductive knowledge, like dependencies between relationships (§ 3), that cannot be represented with DLs. Compositions of relation (Ex15) and relations whose arity is more than 2 are also needed (Ex16), but cannot be represented by these languages.

- Ex15.** **composition** between relations is not provided but required  
 /a concept which has a location, which is included in a region, occupied by another concept C' /  
 isLocatedIn  $\circ$  isSubAreaOf  $\circ$  hasLocation  
 A possible solution for representing composition is using rules.



**Ex16. n-ary relation is not provided but required**

/ternary relation: a sulcus is a separator for two lobes, or two gyri, or one gyrus one lobe/

Separation := AnatomicalConcept  $\wedge$  (= 1 separator Sulcus) (1)

$\wedge$  (= 2 separate Lobe  $\vee$  Gyrus)

parts(S, V)  $\wedge$  1stPart(V, A)  $\wedge$  2ndPart(V, B)  $\rightarrow$  separation(S, A, B) (2)

Frames and description logics allow only binary relations. Possible solutions for representing n-ary relations include relation reification, i.e. to represent it by a concept, e.g. Separation (1), and rules (2) like CARIN-ALN rules [9].

**Ex17. inference rule is not provided but required (#15)**

Rule # 1 : IF A is part of B THEN A has the same side as B

isAnatomicalPartOf (A,B)  $\wedge$  hasSide (B,C)  $\rightarrow$  hasSide (A,C)

Rule#2 : IF a functional activity A is located in a part B of an anatomical structure C, THEN A is also located in the anatomical structure C.

isFunctionallyLocatedIn (A,B)  $\wedge$  isAnatomicalPartOf (B,C)  $\rightarrow$

isFunctionallyLocatedIn (A,C)

Rules are needed to support several tasks: (1) to generate the concepts and/or the relationships that associate them, for the ontology construction and its updates, (2) to check the ontology consistency for knowledge verification, (3) to combine or connect knowledge for managing multiple ontologies, (4) to process queries, or formulate conjunctive queries over multiple heterogeneous sources for an information integration system

**Ontology construction.** There are a lot of dependencies between the multiple concepts of the ontology and their relations. It is difficult, even quite impossible, to create all them or to maintain the ontology by hand. Adding a concept or a relationship may trigger the chaining creation of many other concepts or relationships. For instance, when stating that central sulcus separates the preCentral gyrus and the postCentral gyrus, additional separation relationships are derived, whose corresponding knowledge has to be added in the ontology, that is from the following facts F1, F2, F3, and Rule R1 the conclusion C is derived, and the corresponding knowledge should be added.

**Rule R1:** hasPart(A, B)  $\wedge$   $\neg$  hasPart(B, C)  $\wedge$  separates(S, A, C)  $\rightarrow$  separates(S, B, C)

PreCentralGyrus is an anatomical part of frontalLobe (F1)

frontalLobe and parietalLobe have no common part (F2)

centralSulcus separates precentralGyrus and parietalLobe (F3)

centralSulcus separates frontalLobe and parietalLobe (C)

For example, a single rule such as «if a sulcus S separates two gyri G1, G2 that belong to distinct lobes (G1 is part of L1, G2 is part of L2), then S separates G1 from L2, G2 from L1, and L1 from L2» would generate 221 relations in the brain cortex ontology presented in [4].

Therefore, it is proposed to build by hand a minimal initial set of independent concepts and relationships, to represent all the dependencies by a rules base, then to automatically generate all the knowledge inferred in applying the rule base.

A first experience has been done to test this approach, which needs further investigation. Rules are implemented in Prolog, the domain of variables is the set of AnatomicalConcepts. Each concept of the ontology is represented by a constant e.g. CentralSulcus is represented by the constant c-centralSulcus, relations between

concepts are represented by facts e.g. separates (c-centralSulcus c-precentralGyrus c-parietalLobe (F3). Using this trick to represent the ontology concepts, roles, and rules in the same logical framework, it has been possible to generate new concepts of the ontology from a first set of « independant » concepts. But for that, it has been necessary to translate the knowledge of the ontology already represented in DAML+OIL into Prolog. A unified formalism integrating rules and DL would be more suited. It would prevent from duplicating the structural knowledge already described in the ontology, and from translation errors. The rules KB should represent only the deductive part of knowledge.

**Consistency checking.** When updating the ontology, it should be checked that added concepts and relationships do not conflict with previous ones e.g. a left-lateralized concept cannot be a part of a right-lateralized concept, and conversely. Constraint like the Rule R2 together with the above *Rule # 1*, enable to check such inconsistencies. After saturation, all the inconsistent concepts are explicitly reported.

**Rule R2** :  $\text{hasSide}(X, \text{leftSide}) \wedge \text{hasSide}(X, \text{rightSide}) \rightarrow \perp$

Such a rule base serve as a conceptual model of consistency (similar to the usual approach for validation of knowledge bases), that can be used to check the ontology consistency.

**Managing multiple ontologies.** Ontologies of related domains may have to be combined. For example, an ontology of brain anatomy may have to be merged with an ontology of the anatomy of the whole body. Both may be connected to an ontology of pathologies, in order to provide a disease localization support. Merging, or connecting ontologies requires to handle new dependencies. For instance, a tumor located in the part of an organ is also located in the organ. Such propagation of relations along transitive role as “part-of” are very frequent in medicine. However, all these dependencies cannot be statically and manually generated. Defining rules like R3 expressing the dependency of relations is a solution to manage them dynamically and automatically.

**Rule R3**:  $\text{tumor}(T) \wedge \text{anatomicalConcept}(A) \wedge \text{anatomicalConcept}(B) \wedge \text{hasAnatomicalPart}(A, B) \wedge \text{locatedIn}(T, B) \rightarrow \text{locatedIn}(T, A)$

**Processing or representing queries.** Rules like R3 are required for processing queries such as “retrieve all images showing a tumor in the frontal lobe”. Indeed, it is wanted to obtain all the images showing either (1) any type of tumor in a frontal lobe, (2) tumors in any specialization of frontal lobe e.g. in the left frontal lobe, (3) but also tumors located in any part of a frontal lobe, or (4) any combination of the three previous cases, e.g. “a glioblastom in the left precentral gyrus”, which is a specific tumor located in a specific part of a frontal lobe, has to be found.

These different use cases highlight both the benefits that can be expected from rules for building, managing and insuring ontologies consistency, and the necessity to rely on such explicit rules for managing multiple ontologies and representing or processing queries. Therefore, OWL DL is not sufficient, and a unified formalism integrating rules and description logics, would provide an elegant solution to construct, maintain, combine, and exploit ontologies.

### 5.3 Needs of Metaclass and of Modularity and Reuse Mechanisms

OWL DL might be not sufficient for some applications, and OWL Full required to represent metaclasses that may be needed and are not provided by OWL DL. For instance, metaclasses are a possible means for connecting an ontology like the cortex anatomy ontology with existing medical standards like UMLS™. Since Metaclass exist in Protégé, defining a metaclass with a slot ‘UMLS-ID’ connecting the ontology concepts to the UMLS™ concepts is possible in Protégé, but not in DAML+OIL. They may also become necessary for compatibility when merging several anatomical models, as advocated in [16]. Although it is not impossible, manipulating metaclasses in description logics is not straightforward. However metaclasses are legal in OWL-Full, but with no computational guarantees.

**Ex18. metaclasses** are not provided in DAML+OIL ( and OWL DL) but may be required (#16)

```
/The class FrontalLobe, instance of the metaclass MetaAnatomicalConcept, is related by the
property UMLS-ID to the UMLS Concept Unique Identifier C0016733/
<MetaAnatomicalConcept rdf:ID="FrontalLobe">
<UMLS-ID rdf:resource="&rdfs;Literal">C0016733</UMLS-ID>
```

Modularity and reuse mechanisms, similar to those proposed in software engineering, for modular specifications, are not provided but are also required to import ontologies or to reuse a general ontology, while respecting semantics constraints.

### 5.4 Results

Using PROTEGE 2000 and DAML+OIL for the brain cortex ontology pointed out a number of limitations of these languages and has led to the following conclusions about requirements of expressiveness that a Web ontology language should cover:

- First, representing the brain cortex anatomy ontology led to difficulties with both languages, but many limitations of Protégé are overcome by DAML+OIL, thanks to the enhanced expressiveness of *SHIQ* description logics versus frames.
- Next, it comes out that most DAML+OIL constructors (Table 1) and axioms (Table 2) in particular negation, disjunction, inverse, were needed for the ontology and would certainly be in a Web language for biomedical ontologies
- Equivalence of classes or relations, subclass and subproperty are key axioms to assert relationships between classes and relations of separately developed ontologies, thus are specially required for managing several Web ontologies.
- Finally, the main limitation encountered with DAML+OIL, and that the future Web Ontology Language shall overcome, is the lack of rules (§ 5.2)
- Metaclasses is a possible means for expressing compatibility of ontologies with the existing medical standard terminologies like UMLS™. Modularity and reuse mechanisms are required respectively for assembling elementary ontologies into more complex ones or for reusing more general ontologies

In conclusion, an expressive DL similar to DAML+OIL is suited to express the complex taxonomic knowledge, but rules are also required to enable representing the deductive knowledge necessary for supporting several tasks (ontology construction, maintenance, verification etc.). They might also be useful for representing predicates

of arbitrary arity. Inferential services related to both formalisms are required. Finally, metaclasses might be used to take advantage of the existing medical standard terminologies. However to keep computational guarantees, other solutions shall be found.

## 6 Discussion

W3C standards but also other formal languages are available for Web ontologies. Table 1 and Table 2 compare the main class constructors and axioms supported by Protégé-2000 and DAML+OIL, which is quite similar to OWL-DL, to those of OWL Lite which is less expressive, and of CARIN-ALN [9], an hybrid language combining DL and rules.

From a formal point of view, DAML+OIL (and OWL DL) is quite equivalent to the description logics *SHIQ* extended by the oneOf constructor and datatypes together with a nice set of algebraic axioms. It can make use of the FaCT system which provides a reasoner with sound and complete tableaux algorithms to reason on ontologies, thus supports automatic tasks like ontology consistency checking, concepts classification, instantiation. CARIN-ALN is based on the less expressive ALN description logics, but combines it with a powerful rules language. OntoClass provides for CARIN-ALN the same services as FaCT, but subsumption and satisfiability are polynomial instead of exponential. Moreover, thanks to its rules, CARIN-ALN can serve as a query language to integrate heterogeneous sources via mediators built with PICSEL [9] (were queries reformulation is decidable).

The above study leads to conclude that ideally an hybrid language integrating an expressive DL with rules, similar to CARIN-ALN or TRIPLE [24] would benefit to medical ontologies. Besides, it might serve as a query language in order to search medical information on the Web and to query heterogeneous sources. But, combining description logics with rules implies to restrict, either the description logics part or/and the form of rules, to remain decidable and to have sound and complete algorithms [12]. An open question is to define a relevant subclass of DL and of rules to be integrated into a uniform language, suited to represent medical ontologies.

**Table 1.** Main class constructors (used in the brain cortex ontology)

| Constructor    | Protégé 2000       | DAML-OIL (DL syntax)                      | Example   | OWL Lite  | CARIN-ALN                   |
|----------------|--------------------|---|-----------|---|-----------------------------|
| 1. conjunction | multiple hierarchy | $C1 \wedge C2$                            | Ex4       | $C1 \wedge C2$ , for named C1, C2                               | $C1 \wedge C2$              |
| 2. disjunction | No                 | $C1 \vee C2$                              | Ex2       | No  | No                          |
| 3. universal   | Yes                | $\forall r.C$                             | Ex1       | $\forall r.C$   | $\forall r.C$               |
| 4. existential | No                 | $\exists r.C$                             | not used. | $\exists r.C$   | No                          |
| 5. negation    | No                 | $\neg C$                                  | Ex7       | No  | $\neg C$<br>for C primitive |
| 6. cardinality | single or multiple | $\leq n r C$<br>$\geq n r C$<br>$= n r C$ | Ex14      | $\leq n r C$<br>$\geq n r C$<br>$= n r C$<br>for $n = 0$ or $1$ | $\leq n r$<br>$\geq n r$    |

**Table 2.** Main axioms (used in the brain cortex ontology ontology)

| Axiom                    | Protégé 2000 | DAML-OIL (DL syntax)  | Example | OWL Lite                                | CARIN-ALN  |
|--------------------------|--------------|---|---------|---|--|
| 7. subsumption           | Yes          | subClassOf<br>$C1 \subseteq C2$   | Ex10    | subClassOf<br>$C1 \subseteq C2$         | $C1 \subseteq C2$<br>$C1$ primitive                  |
| 8. class equivalence     | No           | sameClassAs<br>$C1 \equiv C2$   | Ex9     | sameClassAs<br>$C1 \equiv C2$           | No   |
| 9. exclusion             | No           | disjointWith<br>$C1 \subseteq \neg C2$  | Ex5     | No                                      | $C1 \wedge C2 \subseteq \perp$<br>$C1, C2$ primitive |
| 10. disjoint union       | No           | disjointUnionOf<br>f<br>$C \equiv C1 \vee C2$<br>$C1 \wedge C2 \subseteq \perp$ | Ex8     | No                                      | No   |
| 11. subproperty          | Yes          | subPropertyOf<br>$r1 \subseteq r2$  | Ex10    | subPropertyOf<br>f<br>$r1 \subseteq r2$ | No   |
| 12. property equivalence | No           | samePropertyAs<br>s<br>$r1 \equiv r2$   | Ex13    | samePropertyAs<br>s                     | No   |
| 13. inverse              | Yes          | InverseOf<br>$r1 \equiv r2^{-1}$  | Ex12    | inverseOf                               | No   |
| 14. transitivity         | No           | transitive  | Ex11    | transitiveProperty                      | No   |
| <b>15. Rule</b>          | No           | No  | Ex17    | No                                      | Carin-rule   |
| <b>16. Metaclass</b>     | Yes          | No  | Ex18    | No                                      | No   |

This study about expressiveness is a first step, to go further it should be useful that the main usage of a Semantic Web expected by the biomedical community are more precisely specified and its strong requirements stated: are decidability, sound and complete algorithms, efficient reasoning, essential or not?

## 7 Conclusion

Most people nowadays agree that a Web ontology language shall have formal semantics. We believe that it shall be enough expressive to allow a fine and precise representation of both terminological and deductive complex knowledge. But it has also to provide efficient means to reason with huge ontologies: (1) automatic ontology classification and consistency checking, and also (2) services for querying heterogeneous and distributed sources, (3) modularity and reuse mechanisms for assembling elementary ontologies into more complex ones or reusing more general ontologies. Thus, OWL is a good candidate for the taxonomic part, but it is not sufficient and should be extended by rules for representing the deductive component of knowledge. However, the combination of an expressive DL e.g. *ALNR* with rules as Datalog, enlarges the search space. The challenge is to identify an hybrid formalism combining a subclass of OWL with rules, that allows to remain decidable and to have sound and complete algorithms for subsumption and satisfiability, and also if possible good properties for the reformulation of queries upon heterogeneous information sources. A trade-off shall be found in restricting, either OWL to an

appropriate sublanguage or/and the form of rules to appropriate versions of Datalog/RuleML. OWL Full with the syntactic freedom of RDF allows metaclasses, but offers no computational guarantees. Another important challenge, particular important for the biomedical domain, is to find suited solutions for relating formal ontologies to existing domain thesauri, that cannot be ignored, and for modular specification of Web ontologies.

## References

1. Bechhofer S., Horrocks I., Goble C., Stevens R. OILED: a Reason-able Ontology Editor for the Semantic Web. Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, Vienna. Springer LNAI Vol. 2174, (2001) 396–408
2. Bowden, DM and Martin, RF. NeuroNames Brain Hierarchy, *Neuroimage*,2 (1995) 63–83
3. Brinkley J.F. and Rosse C. Imaging informatics and the Human Brain Project : the role of structure, *Yearbook of Medical Informatics* (2002) 131–148
4. Dameron O., Burgun A., Morandi X., Gibaud B. Modelling dependencies between relations to insure consistency of a cerebral cortex anatomy knowledge base. Proceedings of Medical Informatics in Europe (2003)
5. DAML+OIL Reference Description. Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. W3C Note 18 December 2001. <http://www.w3.org/TR/daml+oil-reference>.
6. Extensible Markup Language (XML) 1.0 (Second Edition). Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, and Eve Maler, eds. (2000). <http://www.w3.org/TR/REC-xml>.
7. Fellbaum C, ed. WordNet: an electronic lexical database. Cambridge, MIT Press (1998)
8. Fensel D., van Harmelen F., Horrocks I., McGuinness D.L., and Patel-Schneider P. F. OIL An ontology infrastructure for the semantic web. *IEEE Intel. Systems*, 16(2) 38–45 (2001)
9. Goasdoue, F., Lattes, V., Rousset, M.C. The Use Of Carin Language and Algorithms for Information Integration: The PICSEL System *International Journal of Cooperative Information Systems*, 9(4): 383–401, 2000.
10. Gomez-Perez A., Corcho O., Ontology languages for the Semantic Web, *IEEE Intelligent Systems* 17, 4 54–60, 2002.
11. Hendler, J., McGuinness, D.L. The DARPA Agent Markup Language. *IEEE Intelligent Systems* 16(6) 67–73, 2000.
12. Levy A. Y, Rousset MC, The Limits on Combining Recursive Horn Rules with Description Logics, *AAAI/IAAI*, Vol. 1, 1996.
13. Li Q, Shilane P, Noy NF, Musen MA Ontology acquisition from on-line knowledge sources. *Proc. AMIA Symp.* 497–501, 2000.
14. Lindberg D.A, Humphreys, B.L. McCray AT. The Unified Medical Language System. *Meth. Inf Med Aug*; 32(4) (1993) 281–91
15. Noy N. F. Sintek M, Decker S., Crubezy M, Ferguson R. W., & Musen M. A.. Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems* 16(2): 60–71, 2001. <http://protege.stanford.edu>
16. Noy NF, Musen MA, Mejino JL Jr., Rosse C. Pushing the Envelope: Challenges in a Frame-Based Representation of Human Anatomy. Technical Report, (2002). ([http://smi-web.stanford.edu/pubs/SMI\\_Abstracts/SMI-2002-0925.html](http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-2002-0925.html))
17. Ono M, Kubik, S and Abernathey, Geog Thieme Verlag, Atlas of the Cerebral Sulci, Thieme Medical Publishers Inc., (1990).
18. OWL Web Ontology Language 1.0 Reference. Mike Dean, Dan Connolly, Frank van Harmelen, James Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. W3C Working Draft 12 November 2002 <http://www.w3.org/TR/owl-ref/>. OWL Web Ontology Overview :W3C Working Draft 4 March 2003

19. RDF/XML Syntax Specification (Revised) Dave Beckett, ed. W3C Working Draft 23 January 2003. Latest version is available at <http://www.w3.org/TR/rdf-syntax-grammar/>.
20. Rector A. Analysis of propagation along transitive roles: Formalisation of the GALEN experience with Medical Ontologies, 2002 International Workshop on Description Logics DL2002, Toulouse, France, April 19–21, (2002) .
21. Rector. A., Nowlan W.A. and the GALEN Consortium, The GALEN Project Computer Methods and Programs in Biomedicine, 45, 75–78, 1993.
22. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the digital anatomist symbolic knowledge base. *J Am Med Inform Assoc.* 1998 Jan–Feb; 5(1):17–40.
23. Sakai Y. Metadata for evidence based medicine resources, Proc. Int'l Conf. On Dublin Core and Metadata Applications 2001, 81–85, 2001.
24. Sintek M., Decker S. TRIPLE An RDF Query, Inference, and Transformation Language. DDLP'2001, Japan, (2001).
25. Sure Y., Erdmann M., Angele J., Staab S., Studer R. and Wenke D. OntoEdit: Collaborative Ontology Engineering for the Semantic Web. In Proc. of the first Inter. Semantic Web Conference 2002, June 9–12 2002, Sardinia, Italy, (2002).
26. Toga A.W. Neuroimage databases : the good, the bad and the ugly, *Nature reviews neuroscience* vol. 3 302–309, 2002.
27. Varzi A. Parts, Wholes, and Part-Wholes Relations: the Prospects of Mereotopology. *Data and Knowledge Engineering*, 259–286, 1996.