OntoNeuroBase: a Multi-Layered Application Ontology in Neuroimaging

Lynda TEMAL\textsuperscript{a}, Pascal LANDO\textsuperscript{b}, Bernard GIBAUD\textsuperscript{a}, Michel DOJAT\textsuperscript{c}
Gilles KASSEL\textsuperscript{b}, Anne LAPUJADE\textsuperscript{b}

\textsuperscript{a}Unit / Project Visages U746, INSERM/INRIA/CNRS/University of Rennes 1, Rennes, France
E-Mail: \{lynda.temal, bernard.gibaud\}@irisa.fr

\textsuperscript{b}LaRIA, CNRS (FRE 2733) and Jules Verne University of Picardie, Amiens, France

\textsuperscript{c}INSERM UM594, Grenoble, France
E-Mail: michel.dojat@ujf-grenoble.fr

Abstract. OntoNeuroBase is an application ontology that is being developed within the NeuroBase project, which seeks to create a federated system for the management of distributed and heterogeneous information sources in neuroimaging. Having adopted a specific, multi-layered, modular approach to ontology design, we used DOLCE as a foundational ontology together with three core ontologies: I&DA for modelling documents (texts and images), COPS for modelling programs and software and OntoKADS for modelling problem solving activities. Here, we report on how we built OntoNeuroBase by refining the concepts present in the existing modules. Neuroimaging is a very active and rapidly changing field. It is essential to ensure that a newly developed ontology is compatible with other available ontologies and to enable extension of the new ontology to a variety of neuroscience applications. The work reported here is in line with these ambitious objectives.

Introduction

We are currently working on the construction of an application ontology in the context of the NeuroBase\textsuperscript{1} project for sharing and reuse of data and tools in neuroimaging. This paper addresses the methodology used for such design.

NeuroBase (the project)

Today, neuroscientists are able to explore brain function and dysfunction using variety of neuroimaging techniques (e.g. Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) and Magnetoencephalography (MEG)). Furthermore, the mapping of functional activations with anatomical brain structures is based on sophisticated image processing techniques (such as segmentation, registration and multi-modality fusion) and use specific statistical methods for longitudinal data analysis. Two major difficulties have emerged in neuroimaging information management: the huge quantities of data produced (around 1 Gb per subject) and the reuse and

\textsuperscript{1}http://www.irisa.fr/visages/demo/Neurobase/index.html
sharing of data/programs that have been acquired/developed in different centres - typically clinical. In order to solve these problems, a number of ongoing projects have suggested the use of a centralized data base (Civet\(^2\), the Mind Institute\(^3\) projects), standardized computer tool development (National Alliance for Medical Image Computing\(^4\) project) or of a customized, distributed environment via grid architectures (UK-escience\(^5\), BIRN\(^6\) projects).

The NeuroBase project has adopted a different approach by promoting a federated system for the management of distributed and heterogeneous information sources in neuroimaging (given that these sources are located in a range of different centres typically clinical departments in neurology and radiology, and research labs in cognitive neuroscience). The goal is for users to circulate, exchange or retrieve distant neuroimaging information almost as easily as if it were stored locally. This process involves sharing two types of information: neuroimaging data (typically generated in neuroimaging experiments) and image processing programs applied to data present in the distributed system.

An important aspect of this project is the definition of a common semantic model which integrates the main concepts shared by all the partners. Each site participating in the federated system can map its own concepts, data and image processing programs against this repository. We rely on a mediator/wrapper approach in which a mediator offers a central view of all information sources and the associated source-specific wrappers mask the sources’ heterogeneity. The mediator uses the appropriate wrappers to redefine the user query into source-dependent queries. It then recomposes the various responses and formats the final response sent to the user. The semantic repository was built using an ontological approach.

**OntoNeuroBase (the ontology)**

OntoNeuroBase is the application ontology currently being developed within the NeuroBase project in order to enable the sharing of neuroimaging data and image processing tools. It was developed to meet several different objectives: 1) integrating conceptualizations from many different fields (e.g. neuroanatomy, neurophysiology, neuropathology) into a consistent whole; 2) ensuring compatibility with other available ontologies; 3) designing a reference ontology for a broad community of users (currently targeted applications concern epilepsy, visual cortex exploration and Alzheimer’s disease) and 4) designing an easily extensible ontology.

In order to fulfill these objectives, the chosen design approach consisted in structuring the ontology into several clearly identified modules situated at different levels of abstraction (Fig. 1). More precisely, this approach aims at mastering two complexities:

- **conceptual complexity** - it is important to allow for the modelling of complex objects (such as medical images and their processing tools) at different levels of abstraction.

- **design complexity** - two objectives are considered: on one hand, the reuse of modules that have been used and evaluated in the development of other application ontologies and, on the other hand, the ability to work in a distributed fashion for the design of new modules.

In this paper, we present our use of a multi-layered approach, together with the structure of the OntoNeuroBase kernel. Section 1 presents the most abstract reused modules: DOLCE

\(^2\)http://wiki.bic.mni.mcgill.ca/index.php/CIVET
\(^3\)http://www.themindinstitute.org/aboutus.html
\(^4\)http://www.na-mic.org/index.htm
\(^5\)http://www.ixi.org.uk/
\(^6\)http://www.nbirn.net/
(Descriptive Ontology for Linguistic and Cognitive Engineering), as a foundational ontology [1]; I&DA (Information and Discourse Acts), a core ontology initially built for classifying documents as a function of their content [2]; and OntoKADS, a core ontology for the modelling of problem-solving activities [3]. Following sections deal with the extensions brought to the previous ontologies in order to build OntoNeuroBase. Section 2 presents COPS, a Core Ontology of Programs and Software that extends the previous ontologies. Section 3 presents two domain ontologies, one for the image processing tool domain (extending COPS), and the other for the images themselves and their content (extending I&DA). Section 4 assembles comparison elements concerning current ontological efforts related to OntoNeuroBase, and more particularly to our new-built ontologies. Finally, development perspectives for our project are evoked in section 5.

Two separate expressions of OntoNeuroBase operate in our project: a semi-informal OntoSpec [4] version and a formal OWL-DL version, with the former serving as documentation for the latter. Given that our aim in this paper is to emphasize the structure of the ontology, we shall disregard syntactical aspects and merely show graphic subsumption links.

![Figure 1: Main ontologies composing OntoNeuroBase kernel. A descending link between two ontologies O₁ and O₂ means that concepts and relations of O₂ specialize concepts and relations of O₁.](image)

### 1 The Ontological Reference Framework

In this section, we briefly review the main concepts of the abstract modules we reuse.

#### 1.1 Particulars (DOLCE)

In contrast to *universals, particulars* are entities that cannot have instances. DOLCE [1] decomposes the domain of *Particulars* into four separate sub-domains, on the basis of the entities’ different modes of existence (Fig. 2):

- Firstly, one assumes the existence of *concrete entities* that extend in space-time. Of these, and according to a 3D ontological option, entities that "are in time" (*Endurants*) are contrasted with entities that "occur in time" (*Perdurants*), with the former participating in the latter, at some given moment.

- *Endurants* and *Perdurants* are characterized by inherent properties (*Qualities*) that take a "value" within quality spaces (*Abstracts*).
Within *Endurants*, DOLCE further distinguishes between *Physical* and *Non-Physical Endurants* according to whether the entity has direct spatial qualities. We shall see below that this turned out to be fundamental for structuring OntoNeuroBase.

![Diagram of DOLCE's top-level taxonomy](image)

**Figure 2: An excerpt from DOLCE’s top-level taxonomy**

1.2 *Documents (I&DA)*

I&DA is a core ontology in the domain of semiotics that was initially built for classifying documents by their contents [2]. I&DA extends DOLCE by introducing three main types of entities (Fig. 3):

- **Inscriptions** (e.g. written texts, images) are knowledge forms materialized by a substance (e.g. ink, an electronic field) and inscribed on a physical support (e.g. a piece of paper, a hard disk). The peculiarity of these *Physical Endurants* lies in their intentional nature: *Inscriptions* stand for other entities.

- **Expressions** (e.g., texts, equations) are Non-Physical knowledge forms ordered by a communication language. *Inscriptions realize Expressions*, and like *Inscriptions*, *Expressions* are intentional entities conveying contents for *Agents*.

- **Conceptualizations** consist of the means whereby *Agents* can reason about a world. Functionally, one distinguishes between two kinds of *Conceptualizations*: *Propositions*, as a means of describing a state of affairs, and *Concepts*, as a means of classifying entities. *Conceptualizations* can be *expressed* by *Expressions* and *physicallyRealized* by *Inscriptions*.

1.3 *Role modelling*

In this paragraph, we present an additional component of OntoNeuroBase, namely a sub-ontology of "participant roles". Above all, however, we wish to highlight a general principle used to model concepts (such as *Subject*) which we call "relational roles" or simply "roles".

Following the most widespread practice in conceptual and object-oriented modelling [5] (see [6] for general discussion about these concepts), we consider *Roles* as being:

---

8 In the rest of the paper, relation names will be written using a Java-like notation.
Figure 3: I&DA’s top-level taxonomy

- **Anti-rigid** (or dynamic). In fact, even though every Subject is a Person, a Person is only contingently a Subject. In other words, a Person does not necessarily have to be a Subject.

- **Relationally dependent.** Being a Subject requires the existence of an Experiment in which the Subject participates.

As Guarino and Welty have shown [7], this type of characterization can be formally defined by means of meta-properties. For the sake of this paper, we consider three such meta-properties defined by these authors as follows:

- A Role is an anti-rigid and externally dependent concept.

- A Formal Role is a Role which does not carry any identity criterion. This type of Role is not constrained in terms of the nature of its instances. These Roles include participant roles (e.g. Agent, Patient) which define ways for Endurants to participate in Perdurants (Fig. 4a).

- A Material Role is a Role carrying an identity criterion. This type of Role (e.g. Subject) encapsulates both a Formal Role (e.g. Patient in an Experiment) and the type of instance playing this Role (e.g. Person) (Fig. 4b).

Figure 4: Participant roles (4a); Definition of Material Roles by specialization of Formal Roles (4b)

---

8For reasons of space, we cannot present the formal definition of these notions. Interested reader should consult reference [7].


1.4 Reasonings and "Knowledge Roles" (OntoKADS)

The OntoKADS ontology (an extension of DOLCE and I&DA) was designed to build problem-solving models with the CommonKADS methodology [3]. Its contribution to OntoNeuroBase consists mainly of a sub-ontology of Reasonings used for defining how the image content is processed.

In OntoKADS, Reasonings are situated within an ontology of Actions structured according to the following principles (Fig. 5a):

- An Action is a transformation of a world performed by an Agent. Actions are specialized according to the kind of world (either physical or non-physical) in which the transformation occurs.

- A Doing is a transformation of the physical world (e.g. repairing an engine, manufacturing a car).

- A Reasoning is a transformation of the non-physical world (e.g. establishing a diagnosis, planning an experiment). Reasonings correspond to Actions on the physical world that are born "in mind": entities playing the roles of Data and Results of these Actions are descriptions (Conceptualizations) which refer to entities in the physical world.

Taking Reasonings into account prompts one to define specialized roles (Fig. 5b):

- Formal Roles (e.g. Data, Result) are classified into "Formal Knowledge Roles" (e.g. Data To Be explained, Diagnosis Result) which cite a particular Reasoning.

- In turn, these Formal Knowledge Roles allow to define "Material Knowledge Roles" including the type of their instances (e.g. Complaint To Be Explained, Diagnosis Hypothesis), following the general principle presented in section 1.3.

---

In OntoNeuroBase, we have chosen to keep the term "reasoning" used in OntoKADS to denote problem-solving activities, though the term "conceptual action" would be more appropriate.

---

Figure 5: OntoKADS ontology of Actions (5a); Definition of Material Knowledge Roles (MKR) by specialization of Formal Knowledge Roles (FKR)
2 A Core Ontology of Programs and Software

In order to capture the image processing tools which must be taken into account by OntoNeuroBase, we designed a component conceptualizing the general domain of programs and software. This component (named COPS for “Core Ontology of Programs and Software”) extends I&DA by defining the notion of "program" and extends DOLCE, endowed with a notion of "collection”, by defining the notion of ”software”. Lastly, OntoKADS helps by defining the functionality of programs and software. In this section, we present the most abstract concepts of COPS and illustrate them with pedagogical examples.

At a first modelling level, the *Endurant vs Perdurant* distinction enables us to discriminate between a program (as an *Endurant*) and execution of a program by a computer (as a *Perdurant*). By focusing on the nature of the *Endurant*, the distinctions made in I&DA (*Inscriptions*/Express*ions*/Conceptualizations*) prompt us to identify various entities linked to the notion of ”program” (thus shedding light on the term’s polysemy). In fact, I&DA leads to distinguish between the following items (Fig. 6a):

- **Files** as *Inscriptions* inscribed on computer support (e.g. optical disks, main memory, recording tapes). But these *Files* are only one kind of program *Inscriptions*: indeed, paper prints outs or on-screen displays of programs are also program *Inscriptions*.

- **Expressions** specified by means of a programming language (e.g. C++Expression, Java-Expression, etc.). These *Expressions* include *Programs* that *implement* (a sub-relation of *express*) *Algorithms*.

- **Calculus schemata**, or *Algorithms* and *Data Structures* as *Conceptualizations*.

Another *Endurant* entity linked to *Programs* is software. The latter is commonly defined as a ”set” of programs. In COPS, we account for this aspect by relying on a general notion of ”collection” as characterized in [8] and defined in DOLCE as a kind of *Social Object* (Fig. 6b). Our modelling choice is intuitively justified by the following properties that we consider as being attached to *Software*: *Programs* assembled into *Software* may be specified in different programming languages and may vary in both number and nature over time (following an addition or a withdrawal) without altering the identity of *Software*. *Programs* constituting *Software* at a given moment contribute to the functionality offered by the *Software* to its users (unity criterion). The term ”functionality” must be understood here in a broad sense as covering the *Software’s* usage conditions. In this respect, we consider that documents other than *Programs* (e.g. a licence, a user manual) are also constituents of *Software*.

In order to render account of *Perdurants* in which *Programs* are involved, COPS considers the *Actions* (*Reasonings*) that *Programs* allow to carry out: the *allowsToCarryOut* relationship is used to model the functionality of *Programs* and *Software*. Entities participating in these *Actions* as *Data* and *Results* are either *Conceptualizations* or *Expressions*, as in the case of *Programs* exploiting other *Programs*. An example of this latter situation is the compilation of *Programs*. In accordance with the general principle of role modelling, the *Action* concept called *Compiling* allows one to successively define the formal knowledge roles *Compilation Data* and *Compilation Result* and the material knowledge roles *Source Code* and *Runtime*.

3 The Neuroimaging Application Ontologies

Neuroimaging is a broad domain that involves various types of data obtained with different acquisition equipment and processed by different software tools. In this section, we present the ontologies (built from previously described modules) used to handle image processing tools and medical imaging data.
3.1 Image Processing Tools

The Image Processing Tools ontology (IPTO) aims at organizing the tools used for image processing according to the kind of operation that they perform. These tools are *Programs* that may be part of *Software* (e.g. Brainvisa, Matlab, SPM, Mricro or ITK&VTK). Let’s note however that, the expression “image processing tool” is slightly misleading: in OntoNeuroBase, we are more interested in how the image content (that we call “Dataset”) is transformed, rather than in the image itself as a physical rendering of the data. IPTO distinguishes between different kinds of *Processing Tools* (Fig. 7a) according to their functionality modeled by the relationship (allowsToCarryOut) with the corresponding kinds of *Data Processing* (Reasoning) (Fig. 7b). Moreover, to each *Processing Tool* are attached input/output constraints modeled by the acceptsAsData and providesAsResult relationships. For instance, a *Segmentation Tool* acceptsAsData only image data that have been reconstructed.

Unsurprisingly, *Reconstruction Tools* perform the reconstruction of non-reconstructed image data (e.g. projection images acquired using SPECT imaging equipment, MRI raw data). *Registration Tools* are used to register (i.e. to align with respect to a common spatial reference system) two image datasets (for instance an MRI and a PET dataset); the principal result of this type of data processing is a geometrical transformation. *Segmentation Tools* allow the delineation of regions within images, in order to depict a particular anatomical structure or physiological process. *Statistical Tools* produce statistical information from one or several
datasets obtained from one or more subjects (e.g. functional brain maps derived from MRI data). Averaging Tools are used to build templates for inter-subject registration, i.e. datasets used to perform spatial normalization (alignment with respect to a common spatial reference, provided by a specific subject) by averaging image data obtained (using the same kind of imaging equipment) from a population of subjects.

3.2 Images and their Content (Datasets)

The main goal of the Dataset ontology is to define the structure of a "Dataset", i.e. the semantics of the data which compose it and its acquisition or data processing context. The term "Dataset" is widely used by clinicians and scientists in the neuroimaging domain to designate the image and its content as a whole. In our ontology, we use this term to designate the content only. In fact, following the structuration of COPS, the distinctions made in I&DA (Inscriptions/Expressions/Conceptualizations) lead us to identify various entities linked to the notion of "Dataset". Therefore, we distinguish between the following entities (Fig. 8):

- A Dataset as a Proposition, which denotes the content of neuroimaging data. A Dataset is a complex, structured entity which represents data concerning a subject or a group of subjects. It is considered as a description composed on one part of a structured set of values (Set Values) and on an other part of a set of Meta-data. Roughly speaking, Set Values stand for the actual kernel of a Dataset, independently of any encoding format. Meta-data include information which refer to real world entities, and thus documents the following items:
  - the Dataset’s acquisition context (acquisition protocol, acquisition equipment) in terms of calibration and parameter settings (e.g. echo time and inversion time for MRI images, etc.),
  - the anatomical structure or the brain function explored in the scan session (e.g. the left hemisphere, grey matter, vision, audition),
  - the Dataset’s orientation, i.e. the orientation of the subject with respect to the sampled spatial variables.

Furthermore, a Dataset represents an abstract entity, e.g. a function denoting the distribution in space and/or time of a physical quantity such as MRI signal intensity, regional cerebral blood flow or a volume displacement. This aspect is quite complex and cannot be specified in detail within the scope of the present article.

- A Dataset Expression as an Expression, by means of an encoding format (Analyze Dataset Expression, GIS Dataset Expression, DICOM Dataset Expression, etc.).

- An Image as an Inscription (on a computer screen, for example). Images can be further differentiated according to the image dimension (e.g. 2D Image, 3D Image) and the kind of rendering (e.g. Color Image, Black&White Image). Datasets stored in Files represent other kinds of Inscriptions. In turn, File is further differentiated according to the kind of encoding format (e.g. GIS File, Analyze File, DICOM File). These Files realize corresponding Dataset Expressions (e.g. a GIS File realizes a GIS Dataset Expression), and physicallyRealizes a corresponding Dataset.

By focusing on Datasets, we distinguish between some main categories differentiated according to two semantic axes: the first considers categories of Datasets based on the various kinds of Data Processing (e.g. image acquisition, image processing) that produce them; the
second introduces categories based on the kind of structure (anatomical, functional, etc.) being explored.

The first semantic axis categorizes Datasets on the basis of the Reasoning (e.g. Acquiring with MRI, Segmentation Processing) in which they participate as a Result (Acquiring with MRI Result, Segmentation Result) (Fig. 9). Thus, in order to model the roles and the Actions in which Datasets participate, we rely on the ontoKADS ontology (cf. section 1.4). In On-toNeurobase, we distinguish between two kinds of Actions which result in specific kinds of Datasets:

- An Acquiring as a Reasoning, which is a transformation of the non-physical world, in the sense that the action of acquiring a set of images from a physical object (such as the head of the subject) with specific acquisition equipment results in a new Dataset. Acquiring is further differentiated according to the acquisition equipment used (Acquiring with MRI, Acquiring with CT, etc.)

- A Data Processing as a Reasoning, which is a transformation of the non-physical world, in the sense that the processing acts on the image content. The Data Processing is differentiated into a number of principal categories (e.g. Segmentation Processing, Reconstruction Processing) (Fig. 7b) that result in specific kinds of content.

One could argue that Acquiring is a specialization of Doing rather than Reasoning. Our choice is based on the fact that we are more interested in image content (i.e. transformations occurring in the non-physical world) than in the images themselves (i.e. transformations occurring in the physical world).

Thus, Datasets are differentiated according to the kind of Reasoning from which they result (Fig. 9): an Acquired Dataset is the result of Acquiring With specific acquisition equipment, e.g. an MRI Dataset is the result of Acquiring With MRI. A Reconstructed Dataset is the result of Reconstruction Processing, whereas a Non-Reconstructed Dataset has not undergone any Reconstruction Processing. A Segmentation Dataset is the result of Segmentation Processing, whatever the nature of this segmentation (e.g. contour detection, region classification). A Registration Dataset is the result of a Registration Processing; it represents a geometrical transformation (e.g. a 4x4 matrix, a displacement field) of one Dataset into another. A Template Dataset is the result of Averaging Processing; it represents a reference framework for multi-subject image registration.

The second semantic axis: Datasets can be organized according to the structure or function explored (Fig. 9). A distinction is made between Anatomical Datasets that explore brain...
anatomy (e.g. a MR Dataset with a T1 weighted MR signal, which provides good contrast for the various brain tissues), Functional Datasets that explore brain function (e.g. vision, object recognition), Vascular Datasets that explore brain vasculature (e.g. arteries and veins, blood flow) and, lastly, Metabolic Datasets that explore brain metabolism (e.g. MR spectroscopy, showing metabolite distribution).

![Diagram of Dataset top-level taxonomy]

**Figure 9:** An excerpt from the Dataset top-level taxonomy

4 Discussion

In this paper, we have presented three new-built ontologies designed for the NeuroBase project: COPS, the Dataset ontology and the Image Processing Tools ontology. The latter two constitute the kernel of the OntoNeuroBase ontology. We have also emphasized the modular, multi-layered approach adopted for development of OntoNeuroBase. We consider modularity and structuration as two key points for the development, sharing and reuse of ontologies. Modularity is advocated by authors such as Alan Rector [9] for the design of large ontologies where a number of important issues must be dealt with: re-use of modules that are independently modified over time, introduction of new modules and maintenance of existing modules. Our multi-layered approach introduces graded levels of abstraction in order to structure domain ontologies based on core ontologies [10].

This approach notably prompted us to design COPS, a Core Ontology of Programs and Software. Our work must be contrasted with the efforts currently being undertaken to define the WSMO ontology [11], dedicated to the description of web services. In WSMO, general properties of web services as programs (e.g. functional properties) are mixed with specific properties of these same programs (e.g. the error rate generated by the web service). The approach that mostly closely resembles COPS is COS, a Core Ontology of Software defined by Oberle et al. [12]. At first sight, COS relies on different modelling choices. In particular, there is no distinction between the notions of ”program” and ”software” and, somewhat surprisingly, the Data concept subsumes the Program\(^{10}\) concept (on the basis that any program can constitute data for another program). Furthermore, COS relies on different core ontologies - DnS (Descriptions and Situations) [14] and DDPO (DOLCE + DnS Plan Ontol-

---

\(^{10}\)According to this argument, the concept Subject (see Fig. 4b) should subsume the concept Person. By formally attributing meta-properties to these concepts, i.e. by considering that the concept Subject is anti-rigid and the concept Person is rigid, the OntoClean methodology [13] prompts one to rule out this type of subsumption link, considering it as logically erroneous. In our view, this mistake in COS is due to the fact that this ontology lacks a sub-ontology of participant roles.
ogy) [15] rather than I&DA and OntoKADS. A more detailed comparison would necessitate prior matching of these core ontologies.

Equally, various efforts to facilitate the sharing of neuroimaging data and processing tools have been reported. BIRN (the Biomedical Informatics Research Network) focuses on the reuse of terminology repositories such as UMLS and NEURONAMES, which are not based on an upper level ontology. The Internet Analysis Tools Registry\footnote{http://www.cma.mgh.harvard.edu/iatr/} has compiled an inventory of existing neuroimaging tools and describes them in term of location, name, version and required resources. This corresponds to management of a simple list, and an ontology is not used to organize the items. We consider that the design of a specific ontology is essential for addressing the neuroimaging field’s specific needs, even though this task is complex and time-consuming. Hence, OntoNeuroBase constitutes a first step in this direction.

5 Conclusion

The definition of a common semantic model for sharing neuroimaging data and processing tools is an ambitious goal - particularly since this field is very broad and rapidly changing. In fact, our approach involves the conceptualization of a wide range of disciplines (such as software engineering, mathematics, image processing, medical imaging) and medical sub-domains (such as anatomy, physiology and pathology). In this context, selection of a suitable methodology enables the reuse of existing, potentially relevant ontologies and guarantees the overall consistency of the application ontology, resulting from an iterative and integrative building process.

The work presented here illustrates the use of such a methodology, based on reuse of a foundational ontology (DOLCE) and construction of a core ontology (COPS) and two domain ontologies for our primary fields of expertise, i.e. software engineering and image processing in medical imaging.

An extension of COPS is currently being undertaken, in order to meet more elaborate user desiderata, such as the need to carry out more complex data processing (e.g. the simultaneous use of several tools). The COPS ontology is therefore been extended in several dimensions, in order to (i) represent these complex data processing steps, (ii) represent sets of tool “pipelines” that can be used to carry out the multiple steps, (iii) take into account the execution of these tools on different software platforms (e.g. grids) and (iv) represent the data’s genealogy.

A lot of work remains to be done, especially in terms of the need to describe image content when referring to anatomical structures or describing physiological or metabolic processes, for example. The medical informatics community has already performed a good deal of work on medical terminology in order to address such needs (e.g. UMLS, FMA, GALEN), although migration of these systems to formal ontologies has yet to be completed. Our general methodology should facilitate the use of these results when they become available.

References


