

Towards an Ontology for Sharing Neuroimaging Data and Processing Tools: Experience Learned from the Development of a Demonstrator

B. Gibaud^{1,2}, M. Dojat^{3,4}, F. Aubry⁵, H. Benali⁶, O. Dameron¹, S. Kinkingnéhun⁶,
J.P. Matsumoto⁷, M. Péligrini-Issac⁶, R. Valabregue², C. Barillot²

¹Laboratoire IDM, Univ. Rennes I, Rennes, France

²IRISA, Equipe VisAGeS, CNRS-INRIA, Rennes, France

³Unité INSERM U594, Grenoble, France

⁴TIMC, Equipe SIC, CNRS-IMAG, Grenoble, France

⁵Unité INSERM U455, Toulouse, France

⁶IFR 49 / CHR La Pitié-Salpêtrière / CEA-SHFJ, Paris / Orsay, France

⁷Projet SMIS, INRIA, Rocquencourt, France

Abstract

The work described here is part of the NeuroBase project dedicated to the sharing of neuroimaging data and processing tools. It focuses on the creation of a shared ontology, suitable for supporting the integration of heterogeneous data and processing tools in neuroimaging. Given the extensive scope of the project, our present goal was limited to highlighting the most salient concepts related to three axes: (1) Subjects, (2) Exploration modalities, and (3) Data and data processing. Detailed modelling was done for a subset of these concepts, directly related to the development of a demonstrator. The work reported here primarily deals with conceptual modelling, which is a first stage towards defining a formal ontology. Nevertheless, modelling was conducted using rigorous principles such as defined by Guarino and Welty for taxonomy. The results presented here are considered by the authors as preliminary, and intended to trigger discussion and feedback from both neuroscientists and image processing specialists.

1. Introduction

Context: Research in the field of neuroimaging is currently very active and is mainly focused on improving understanding of the detailed mechanisms of brain functions under normal and pathologic conditions, and developing new treatments for brain pathologies, especially those related to ageing such as Alzheimer's or Parkinson's diseases. This research involves two major difficulties ensuing from (1) the various levels of analysis involved (biochemical, cellular, anatomical, physiological, behavioural, etc.) and the various sources of information available from the different exploration techniques that can be used (e.g. in humans, electrical measurement in MEG/EEG, perfusion or diffusion imaging and metabolism-based imaging in SPECT, PET, or fMRI), and (2) inter-individual variability. Therefore, one condition for significant progress in neuroimaging is the setting up of inter-disciplinary collaborations, in order to correlate observations at various scales of normal and pathological brain functioning, to build and assess new theories about brain structure and functions [Viewpoint Neuroscience 01, Koslow 02]. To cope with inter-individual variability and ensure that invariants and variability can be characterised with good statistical significance, large populations of subjects need to be studied, which is generally difficult to achieve for any one given centre. Moreover, many research teams have developed tools to process structural and functional neuroimaging data. Thus, it is of primary importance to be able to share such tools, then to evaluate the most efficient ones and circulate them to the neuroscience community at large.

Managing these challenging open issues requires:

- sharing concepts and data among various areas that are complementary in the neuroscience field,
- sharing data produced in different research centres on different subject populations, to analyse or re-analyse them using various tools, and correlate the results with various observations.

- modelling the data processing chains that can be applied to neuroimaging data.

Such needs have been identified and emphasised by many authors [e.g. Mazziotta 95, Kötter 01,], and several projects exist to facilitate sharing and re-use in neuroimaging, such as those conducted by the International Consortium for Brain Mapping (ICMB) [e.g. Mazziotta 01] and the Biomedical Informatics Research Network (BIRN) [Santini 03]. A first possible approach is to collect all data in a central database [Fox 02]. This is the approach adopted by the fMRI Data Center (fMRIDC) consortium [Van Horn 01]. An alternative approach is to set up a federation of databases wanting to share data and processing tools. Mediator-based systems were proposed in the nineties to overcome data heterogeneity, and ensure interoperability of data processing tools in federated systems [see, for instance, Wiederhold 92 and 97 or Xu 03].

The aim of the NeuroBase project, launched in France in 2002, is to share neuroimaging data and data processing tools, according to a federated system approach, based on the mediator-wrapper concept [Barillot 04].

Aims: The work described here is part of the NeuroBase project. It focuses on creating a shared ontology, suitable for supporting the integration of heterogeneous data and processing tools in neuroimaging. Ontologies have received considerable attention in the knowledge engineering community since 1990 [e.g., Gruber 93, Lenat 95, Guarino 95, Noy 97, Chandrasekaran 99]. This trend has been confirmed in recent years due to the need to improve information semantics sharing on the web, which led to the “semantic web” concept [e.g. Ding 02].

The entities shown in our ontology represent the concepts that need to be shared within the federated system. From a functional viewpoint, the NeuroBase system is intended to support:

- (1) query and retrieval of data from several information bases, i.e. raw data (as provided by image acquisition systems), as well as processed data (obtained using a specific image processing tool). Therefore, selection criteria must be defined, which may concern the subjects (e.g. characterising the mental state of the subjects or, if any, the pathology), the exploration technique, the observations made or the result of applying a specific processing tool (e.g. quantification of brain perfusion, metabolism, delineation of anatomical structures or brain functional areas).
- (2) execution of data processing tools “published” by the various constituent sites of the federated system. The system allows the user to select a particular processing tool, input data and processing parameters.

The results presented here are considered by the authors as preliminary, and intended to trigger discussion and feedback, rather than being straightforwardly usable in the neuroscience community. In particular, the authors are aware that, to be truly relevant for supporting the integration of heterogeneous data, an ontology must be a formal ontology (because it is directly involved in automated data processing). The main focus of the work reported here is actually conceptual modelling.

2. Methods

2.1. Scope

The scope of the NeuroBase project is primarily image processing in neuroimaging, for applications in both cognitive research, typically about the mapping of brain functions in humans, and clinical research about widespread pathologies of the human brain. This involves

focusing on the following axes of description in neuroimaging, and highlighting the most salient concepts whose semantics need to be shared:

- **Subjects:** i.e. describing the various characteristics of the subjects, healthy or suffering from a pathology; this includes a detailed description of the pathology and/or subject's cognitive state;
- **Exploration modalities:** i.e. describing the various imaging techniques used to explore brain anatomy or structure (CT, MRI and MRA), function (PET, fMRI, SPECT, etc.), or electrical brain activity (EEG, MEG and depth electrode recordings);
- **Data and data processing:** i.e. representing complex processing schemes, both to characterise processed data (for subsequent retrieval), and share processing tools within the federated system.

Given this broad scope, our present goal was limited to highlighting the most salient concepts related to the three axes mentioned above, and detailed modelling was performed for a subset of these concepts, directly related to the development of a demonstrator for the NeuroBase system [Barillot et al 04]. At this stage, biology in general and genetics (phylogenesis and ontogenesis) fall outside our scope. To demonstrate the soundness of our approach, we applied it to two applications representative of those encountered in neuroimaging experiments, namely the processing of functional images exploring the visual cortex in humans, and the processing of functional images in the pre-operative exploration of patients with drug-resistive epilepsy.

2.2. Methodological approach

General approach: The design of re-usable ontologies is a recent discipline and methodologies for their development are still in their infancy [Gomez-Perez et al. 04]. Therefore, our approach may be qualified as pragmatic since it involves several levels of conceptual knowledge:

- the ontology proposed by the fMRIDC consortium, whose aims and scope clearly fit to a subset of NeuroBase;
- the available standards (i.e. DICOM) defined to facilitate the exchange of medical images and related data: images collected in our databases are obtained using these standard formats and exchange protocols;
- the concepts proposed by researchers involved in this project based on their own experience of medical information modelling and knowledge of medical image processing.

In order to achieve rigorous structuring and reasoning about the ontological nature of the entities within the federated system, our ontology must include concepts such as those met in “upper level ontologies“, such as Cyc, Wordnet or Ontoclean. Our strategy was to rely as much as possible on existing work, rather than develop idiosyncratic concepts, unless we had good reason to do so.

Integration of the various knowledge sources: This is undoubtedly the most difficult aspect of the work, since – from a modelling perspective - these sources make radically different assumptions, and use very different representations. Integration was therefore done “manually”, considering relevance to the ontology’s ultimate goals as a primary criterion (ontology commitment). A second important aspect was to select concepts providing the appropriate genericity level in order to guarantee extensibility. This was felt to be particularly important given the diversity of imaging modalities and image processing tools, which are salient characteristics of the neuroimaging field. An additional difficulty arose from the necessary tradeoff between two contradictory concerns: (1) to design a flexible enough model

to be able to integrate existing neuroimaging data repositories, and (2) to design a strict enough model to support applications such as processing data with a new available tool (assuming that enough details about the images, for instance about pre-processing, are generally available).

Given the diversity of the existing knowledge sources, the modelling process was a combination of top-down and bottom-up analysis, embedded in an iterative process.

The general consistency of the taxonomic relations was assessed using the principles of Guarino and Welty [Guarino 00], based on identification of the meta-properties of the entities: *rigidity* referring to a property “that necessarily holds for all its instances and in every possible world”; *unity* denoting “the problem of distinguishing the parts of an instance from the rest of the world by means of a unifying relation”; *identity* referring to “the problem of distinguishing a specific instance of a certain class from other instances of that class by means of a characteristic property”, and *external dependence* referring to “the need for the existence of some instance of another entity, which is neither part nor constituent of a given entity”. Possible values for each meta-property are listed in the legend of figure 1.

Representation: As previously stated, emphasis was placed on conceptual modelling, which explains our choice of UML (Unified Modeling Language) class diagrams as a specification language. These models were then translated into a relational database schema in SQL. The reason for this choice is that the demonstrator’s application uses SQL as a query language, and expresses the queries using the ontology’s entities.

Validation: Validation will ensue from the development and use of the NeuroBase demonstrator [Barillot 04]. In particular, we will be able to assess the ability of the ontology to map to existing imaging repositories - through data wrappers - and to fit with the users’ views - through the use of the NeuroBase system user interface (definition of queries for cases and image retrieval, or execution of image processing tools).

2.3. Major knowledge sources

fMRIDC ontology: The fMRIDC ontology was issued in 2003. It is based on a former set of information models (entity-relationship diagrams) called NEUROCORE, developed as part of the University of Southern California Brain Project. The fMRIDC ontology is represented in Protégé, an ontology editor developed by the Stanford Medical Informatics group [Musen 98]. It includes a taxonomy of classes, with four upper-level entities: *Event*, *Object*, *Information* and *Reference*. It includes the most salient concepts involved in MRI imaging, organised in a taxonomy: *Scan session*; *Trial*; *Age*, *Handedness*, *Health status*, *Current medications*, *Cognitive status*, *Education* – subsumed by a class *Assessment*; *Research*, *Subject* – subsumed by a class *Person*; *Machine*; *Experimental protocol*, *Scanner protocol*, *Article*, *Experimental Group*, *Keywords*, *Data*, etc. – subsumed by *Information* (via several intermediate concepts such as *Specification* and *Report*).

These entities cover information related to experimental and data acquisition aspects, but not image processing such as segmentation, registration and statistical analysis of images. Moreover, this ontology is primarily focused on MRI, although some concepts are general enough to embrace other imaging modalities.

DICOM: The DICOM standard provides detailed specifications of formats for communicating medical images and waveforms. It is now implemented worldwide by the major manufacturers of imaging systems and PACS. Consequently, most images (CT, MR, PET, and Nuclear Medicine) are now received in DICOM format. The specifications include information models describing a *Patient / Study / Series / Composite objects* (e.g. *Images*,

Presentation states, Structured Reports) hierarchy. The detailed characteristics of each type of Composite object (e.g. MR Image IOD) are provided as a list of modules. A module is a list of related Data Elements. No explicit models exist in the standard to document this information structure. It has to be retrieved from the syntax used in the module presentation. Therefore, DICOM primarily covers the technical aspects of image acquisition, but not aspects such as stimulation paradigms (e.g. in fMRI). Provisions exist in the standard to document observations made on images, through the concept of *Structured Report*. *Structured Reports* are used to represent the interpretations of a radiologist, or an algorithm (e.g. Computer Assisted Detection tools). However, so far these have not been used to document image processing processes in detail.

Experience of researchers: Several members of the NeuroBase consortium have solid experience of image processing in neuroimaging, through the use of publicly available tools, (such as the Statistical Parametric Mapping package (Friston 95) or the development of home-made libraries (such as the VIsTAL library in Rennes or the BrainVisa environment in Orsay and Paris). This experience, as well as the implementations available, provide a valuable input in the design of the ontology.

3. Results

Figure 1 shows an overview of the major entities, organised in a taxonomy. The main concepts involved in the NeuroBase ontology appear in italics.

Most of the upper-level entities are consistent with the fMRIDC ontology. However, several changes were made. First, we introduced a *Process* entity, to subsume *Study* (referred to as *Experiment* in fMRIDC), *Body process* (needed to model physiology and pathology) and *Data processing*. Similarly, we introduced *Anatomical structure* as a subclass of *Object*. We felt the way anatomy, physiology and pathology were modelled in fMRIDC (as subclasses of *Keyword*) to be inappropriate for our purposes. Please note that in our view, both *Anatomical structure* and *Body process* should be seen here as subject-dependent rather than canonical knowledge. *Artefact* subsumes all kinds of equipment, such as physical equipment (e.g. *Acquisition equipment*) or software (e.g. *Processing tool*).

Guarino and Welty's meta-properties were useful for clarifying modelling assumptions and correcting certain misconceptions. For example, we had considered that *Specification*, *Report*, *Data* and *Processing tool* (i.e. software) were subsumed by *Information*, in the sense that all of these consisted of digital data, and therefore *Information* was initially considered $\sim U$ (no instances have a unity condition). It transpired that this approach was inconsistent, because documents carrying a unity condition such as a *Data set* or a *Scientific publication* could not be subsumed by *Information* (a $+U$ property cannot be subsumed by a $\sim U$ one). This led to a slightly different conception, where *Information* is no longer considered as "consisting of digital data" but denotes "any sort of message represented in digital form" (with $+U$). Consequently, *Processing tool* was no longer considered as a subclass of *Information*, but rather a subclass of *Artefact* (as *Acquisition equipment*). Similarly, in a previous version of the ontology, *Experimental Group* (with $\sim U$) was modelled as a subclass of *Information* (with $+U$), as in fMRIDC. The analysis of meta-properties showed that this approach was inconsistent (properties with incompatible unity conditions are disjoint), and therefore *Experimental group* became a subclass of *Group of people*, sharing compatible unity conditions.

The following sections focus on the three major aspects mentioned above: (1) subjects, (2) exploration modalities, and (3) data and data processing tools.

3.1. Subjects

Subject characterisation is partly application-dependent. Moreover, applications related to brain pathology may require a complex description of pathological cases. As a general principle, we suggest making a distinction between observations that are valid for an entire study (*Persistent Pathology Description* and *Persistent Cognitive Description*) and those that concern shorter periods of time such as a single imaging procedure (*Transient Pathology Description* and *Transient Cognitive Description*). Thus, what is actually explored during an imaging procedure is denoted by the investigator in a *State of the Subject* entity, which refers to the transient descriptions relevant to this procedure. We felt this concept to be lacking in the fMRIDC ontology. For example, characterisation of an ictal SPECT (i.e. exploring the patient's state during an epilepsy seizure) is described by a *State of the Subject* entity providing further details about this particular seizure (precise onset time, seizure symptoms and signs, duration, etc.).

3.2. Exploration modalities

The concern here is to characterise the context of neuroimaging data creation. Both acquired data and processed data are created in reference to a *Study* that denotes the goal pursued by the investigator. This goal is important for indexing data. *Study* should be understood here in the sense of “a research project”, rather than a straightforward exploration of a subject (such as in DICOM). A *Study* aims at studying one (or more) *Brain Function*, *Anatomical Structure*, *Physiological Process*, *Pathology* or *State of Subject*. The exploration modalities are described by the entity *Scan Session*, which focuses on a particular neuroimaging exploration. *Behavioural Assessment* subsumes all non-imaging (i.e. neuropsychological) tests aimed at characterising the capacities and cognitive skills of a subject. All information related to the technical aspects of image acquisition can easily be retrieved from the DICOM standard.

3.3. Data and data processing tools

The representation of processed data and processing tools is a far more complex issue. As emphasised previously, in NeuroBase it is approached from two complementary viewpoints, (1) to enable the query and retrieval of processed data, and (2) to enable the execution of shared data processing tools within the federated system. Our approach considers the constraints arising from these two viewpoints. It consists of producing a detailed representation of the processing tools available for their potential use in the federated system, as well as the way the processing tools were used to produce specific processed data.

Therefore, an important distinction is introduced between *Raw Data-Set* and *Processed Data-Set* (see Figure 2). *Raw Data-Set* refers to data actually measured in a *Subject*, for example projection images in SPECT, or FID (free induction decay) data in MRI, whereas *Processed Data-Set* concerns data resulting from the execution of any processing tool. We introduce a classification of four basic processed data types: (1) *Reconstructed Data-Set* represents the result of any reconstruction algorithm applied to raw data, e.g. MRI images, or current dipoles representing activity generators in EEG/MEG; (2) *Processed Reconstructed Data-Set* describes the result of the application of any processing tool to reconstructed data (such as segmentation and unfolding for structural images, or statistical procedures for functional images); (3) *Processed Acquired Data-Set* describes the result of processing acquired data such as filtered signals in MEG/EEG; (4) *Registration Data-Set* describes data (such as registration matrices) resulting from the application of any registration algorithm to two image data sets. All these entities are subsumed by *Data-Set*. A *Data-Set* has a *Container* describing the physical storage of the data, in reference to a particular format (*Format-Descriptor*), assumed to be shared within the federated system.

Application of a processing tool is documented in a *Data Processing* entity (see Figure 3), specifying a *Processing Context* (e.g. date, user), the *Processing Tool* used, as well as the *Data-Set(s)* and *Input Parameter(s)* used as actual input (*Actual Input Definition*), and the *Data-Set(s)* and *Output Value(s)* used as actual output (*Actual Output Definition*). This model is used to represent data processing chains of any arbitrary complexity.

4. Discussion

The integration of heterogeneous data using federated systems may be conceived according to two approaches: a “global as view” model, which builds the common ontology from the concepts existing in the various sites, and a “local as view” model, which is built first, independently of the concepts present in the sites, and then used to map the ontology to the various site entities as appropriate [Rousset 02]. The latter approach was adopted for the NeuroBase project. It follows that the ontology is designed from relatively ill-defined (i.e. non-concrete) material, since existing neuroimaging information sources do not generally include all the information required for successful sharing nor fulfil all the NeuroBase functional requirements. In particular, such information sources are mostly limited to image files, organised in directories whose arrangement reflects their relationship with a common subject or a common study. Most detailed characteristics of subjects and studies are not available in digital form, are incomplete, or not explicitly related to image files. Currently, our specification work includes an in-depth reflection on what information is required and should be available to be (re)usable outside the site that has initially produced it. This explains the relative complexity of our models (compared to what is usually implemented).

Describing subjects raised the problem of representing time-varying phenomena. Our distinction between persistent versus transient descriptions is a pragmatic answer to the need for subject state representation. However, we are aware that this concept is still relative. Some elements that seem persistent in one study may appear transient in another study with a different time scale. For example, a parameter such as seizure frequency is often used as a basic characteristic of epilepsy in a patient. Nevertheless, this parameter is also used as an indicator of success for a surgical procedure (after, for instance, a three-year follow-up period). The representation of complex cases, involving several surgical procedures, clearly demonstrates the limits of our persistent/transient description model.

The modelling of stimulation protocols in functional neuroimaging is also highly challenging. The protocols devised by the investigators to highlight the detailed mechanisms of cognition are of considerable complexity and variety. It is simply unrealistic to conceive a detailed computer model to represent them all accurately. Therefore, our aim was limited to identifying the brain functions and physiological processes to be studied, and simply referring to conventional stimulation protocols such as those listed in the fMRIDC ontology. On the one hand, this ontology was helpful in describing MR acquisition protocols but, on the other hand, it provided no input as regards the processing of MR functional images. Substantial work is still required for the other modalities. Fortunately, nuclear medicine modalities are supported in the DICOM standard, thus information about acquisition parameters is available. The representation of image processing and processed images is more complex. We are not aware of any ontology dedicated to image processing in the field of medical image processing. So, our efforts in this field break new ground. The multiplicity and intrinsic complexity of processing tools (reconstruction, restoration, quantification, segmentation, registration, iconic normalisation, voxel-based morphometry, etc.) make this task particularly challenging. What is presented here should be seen as an attempt to classify and model image processing chains, and requires feedback from the image processing community. The

relationship between image features, such as a region of interest obtained through image processing, and conceptual entities regarding anatomy, physiology or pathology has not been modelled in detail to date. In this regard, the DICOM structured reporting approach is interesting, although it has some limitations, e.g. 3D areas of interest are not supported.

5. Conclusion and prospects

Ontology-based integration of heterogeneous data is still a relatively new research field, although pioneer work on mediation dates back more than ten years. Application to neuroscience has received a fair amount of attention, and is beginning to be implemented in a number of research projects, but, to our knowledge, no operational system actually exists to date [Van Horn 04].

The work presented here is a contribution towards an ontology for supporting the integration of heterogeneous data and processing tools in neuroimaging. Clearly, our preliminary results require consolidating and extending. Our strategy is to confront our models with concrete experience through the setting up and assessment of a demonstrator. Concerning the ontology itself, its extension is envisaged in three directions: (1) in depth (more detailed entities, addition of more specialised entities), (2) in breadth (extension of scope) and (3) in expressivity (using a suitable representation, such as Protégé [Noy 01] or OWL (Web Ontology Language) [Horrocks 03]).

This experience has improved our understanding of the issues raised by the design, use and re-use of ontologies, in fields where few information models exist, or are limited to very narrow subsets of our domain of interest. From this experience, it seems obvious to us that such ontologies can only be built as a result of an iterative process, which requires implementations to be conceived in such a way that extensibility can be guaranteed [Lammari 04].

6. References

- Barillot C. et al., Neurobase: management of distributed and heterogeneous information sources in neuroimaging. DiDaMIC Workshop (this issue), 2004.
- Brinkley J.F. et al., Imaging informatics and the Human Brain Project: the role of structure, *Yearbook of Medical Informatics* 2002, 131-148, 2002.
- Chandrasekaran B. et al., What are ontologies and why do we need them, *IEEE Intelligent Systems* 14(1):20-26, 1999.
- Ding Y. et al., The semantic web: yet another hip ?, *Data & Knowledge Engineering*, 41:205-227, 2002.
- Fox P.T. et al., Mapping context and content, *Nature Reviews/Neuroscience*, 3:319-321, 2002.
- Friston KJ et al. Statistical Parametric Maps in Functional Imaging: a general linear approach, *Human Brain Mapping*. 1995;2:189-210.
- Gruber R., Formal ontology in conceptual analysis and knowledge representation, chapter "Toward principles for the design of ontologies used for knowledge sharing", In *Formal Ontology in Conceptual Analysis and Knowledge Representation*, edited by Nicola Guarino and Roberto Poli, Kluwer Academic Publishers, 1993.
- Gómez-Pérez, A, et al. Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag 2004.
- Guarino N. et al., Formal ontology in conceptual analysis and knowledge representation, Special Issue of the *International Journal of Human and Computer Studies*, 43(5/6):625-640, 1995.
- Guarino N. et al., Ontological analysis of taxonomic relationships, Proceedings of ER-2000: *International Conference on Conceptual Modeling*, LNCS Springer Verlag, 210-224, 2000.
- Horrocks I. et al., From SHIQ and RDF to OWL: the making of a web ontology language, *Journal of Web Semantics*, 1(1):7-26, 2003.
- Koslow S.H., Sharing primary data: a threat or asset to discovery? *Nature Reviews/Neuroscience*, 3:311-313, 2002.
- Kötter R., Neuroscience databases: tools for exploring brain structure-function relationships. *Philosophical Transactions of the Royal Society London*, 356:1111-1120, 2001.

- Lammari N. et al., Building and maintaining ontologies: a set of algorithms, *Data & Knowledge Engineering*, 48:155-176, 2004.
- Lenat D.B., CYC: a large-scale investment in knowledge infrastructure, *Communications of the ACM*, 38:33-41, 1995.
- Mazziotta J.C. et al., A probabilistic atlas of the human brain: theory and rationale for its development, *Neuroimage*, 89-101, 1995.
- Mazziotta J.C. et al., A four-dimensional probabilistic atlas of the human brain, *Journal of the American Medical Informatics Association (JAMIA)*, 8(5):401-430, 2001.
- Musen M.A., Domain Ontologies in Software Engineering: Use of Protégé with the EON Architecture. *Methods of Information in Medicine* 37:540-550, 1998.
- Noy N.F. et al., The state of the art in ontology design: a survey and comparative review. *AI magazine*, 18(3):53-74, 1997.
- Noy N.F. et al., Creating semantic web contents with Protégé2000, *IEEE Intelligent Systems*, 16(2):60-71, 2001.
- Rousset M.C. et al., Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes: le projet PICSEL, *Revue I3 Information – Interaction – Intelligence*, 2(1):9-58, 2002.
- Santini S. et al., The role of internet images in the Biomedical Informatics Research Network, *Proceedings of SPIE*, Vol 5018, Santini S. and Schettini R. (eds), Internet Imaging IV, 2003.
- Toga A.W., Neuroimage databases: The good, the bad and the ugly, *Nature Reviews/Neuroscience*, 3(4):302-309, 2002.
- Van Horn J.D. et al., The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies, *Philosophical Transactions of the Royal Society London*, 356:1323-1339, 2001.
- Viewpoint: Neuroscience, Neuroimaging database, by the Governing Council of the Organization for Human Brain Mapping (OHBM), *Science*, 292:1673-1676, 2001.
- Van Horn J.D. et al., Sharing neuroimaging studies of human cognition, *Nature Neuroscience*, 7:473:481, 2004.
- Wiederhold G., Mediators in the architecture of future information systems, *IEEE Computer* 25(3):38-49, 1992.
- Wiederhold G. et al., The conceptual basis for mediation services, *IEEE Expert* 12(5):38-47, 1997.
- Xu Y. et al., Component-based mediation services for the integration of medical applications, *Artificial Intelligence in Medicine*, 27:283-304, 2003.

Taxonomy

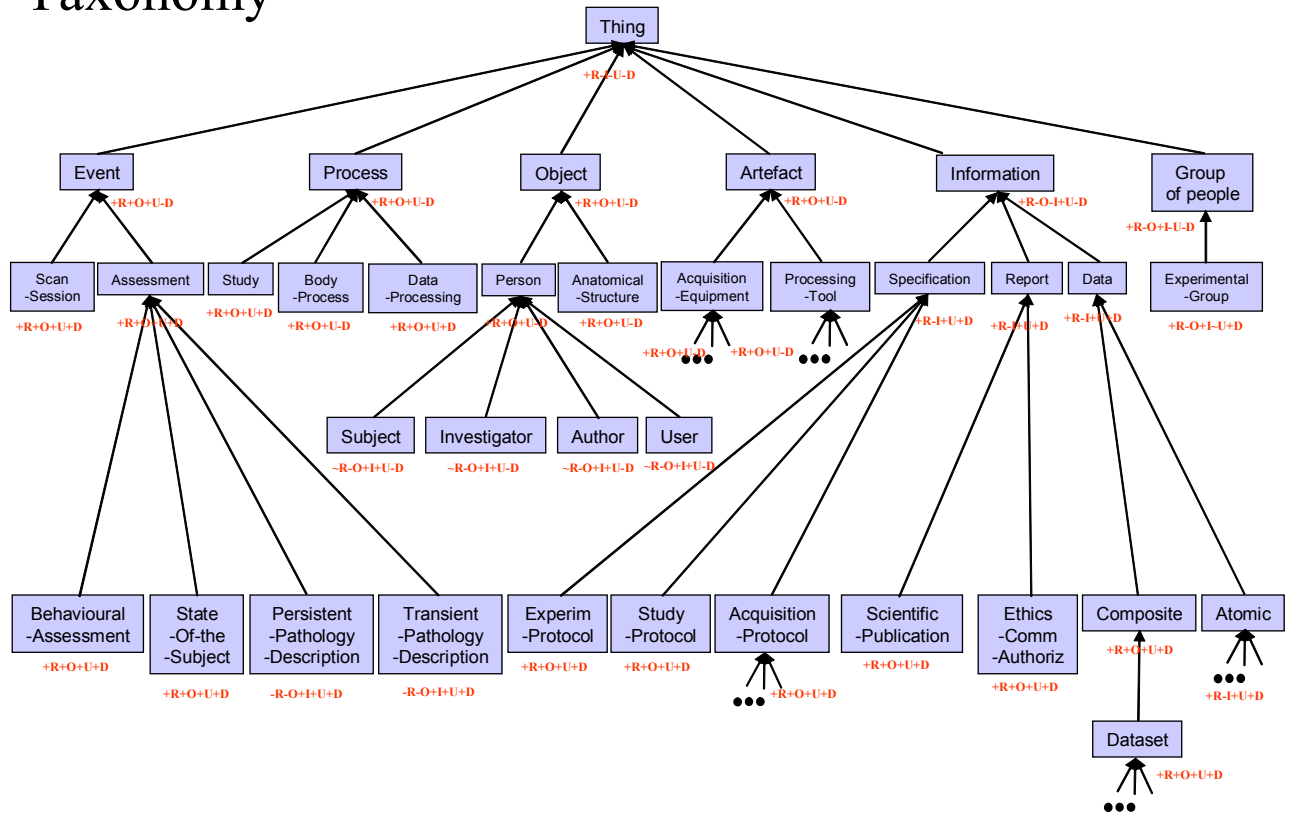


Figure 1: Taxonomy of the major entities of the ontology. Meta-properties are depicted in red.

R refers to *rigidity*: +R (Rigid), -R (Non-Rigid), ~R (Anti-Rigid)

O and I refer to *identity*: +O (entity supplies identity), -O (entity does not supply identity),

+I (entity carries identity), -I (entity does not carry identity)

U refers to *unity*: +U (Unity), -U (Non-Unity), ~U (Anti-Unity)

D refers to *dependence*: +D (externally dependent), -D (Non externally dependent)

Data-Set

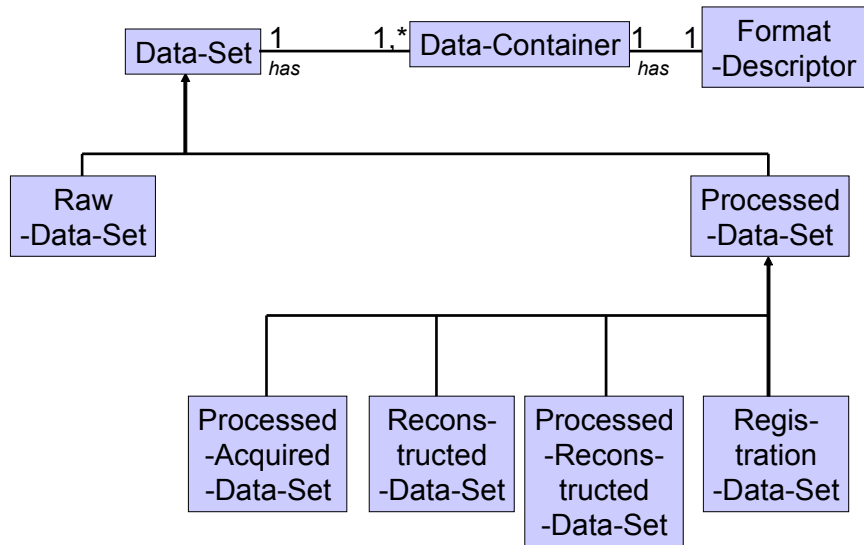


Figure 2: The specialization tree of the *Data-Set* entity

Data-Processing

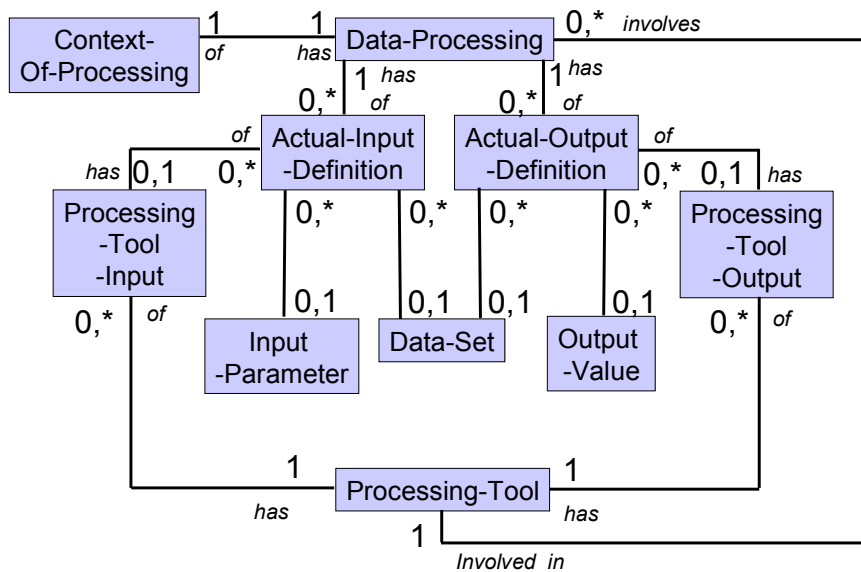


Figure 3: *Data Processing*: relation between *Processing*, *Tool* and definition of actual input and output values