Semantic Representation of Neuroimaging Observations: Proof of Concept based on the VASARI Terminology

Emna Amdouni[1,2] and Bernard Gibaud[1,2]
[1]*E-health department, B-com Institute of Research and Technology, Rennes, France*
[2]*LTSI Inserm 1099, Université de Rennes 1, Rennes, France*

Keywords: Medical Domain Ontology, Knowledge Representation, Ontology Reuse and Alignment.

Abstract: The main objective of this work is to facilitate the identification, sharing and reasoning about cerebral tumors observations via the formalization of their semantic meanings in order to facilitate their exploitation in both the clinical practice and research. We have focused our analysis on the VASARI terminology as a proof of concept, but we are convinced that our work can be useful in other biomedical imaging contexts. In this paper, we propose (1) a methodology, a domain ontology and an annotation tool for providing unambiguous formal definitions of neuroimaging data, (2) an experimental work on the REMBRANDT dataset to demonstrate the added value of our work over existing methods, namely DICOM SR and the AIM model.

1 BACKGROUND AND SIGNIFICANCE

In literature, an ontology is defined as "a formal and an explicit specification of a shared conceptualization" (Gruber, 1993). Ontologies define the formal semantics of vocabularies by specifying axioms, expressed in a logic-based language, that constrain and structure relationships between terms. The main purpose of ontologies is to enable knowledge integration and semantic data querying. In the medical field, semantic web technologies are used to standardize, formalize and share the medical data (coming form both the clinical and research context) is very important (Scheuermann et al., 2009; Seifert et al., 2010; Oberkampf et al., 2012). In this paper, we focused our interest on the domain of cerebral tumors.

In oncology clinical practice, neuroimaging features/phenotypes play an important role; in particular they help clinicians in making their diagnosis, selecting the appropriate treatment and monitoring the therapeutic response to an intervention as for example the Response Evaluation Criteria in Solid Tumors (RECIST) (Eisenhauer et al., 2009). Many radiology-pathology correlation studies have been conducted on cerebral tumors and show that some neuroimaging features are associated to genetic alterations and gene expression (Gutman et al., 2013; Grams et al., 2014). Therefore, suitable management of neuroimaging phenotypes is needed to facilitate their use and reuse in multiple studies regarding imaging biomarkers (Levy et al., 2012; Möller et al., 2009; ESR, 2011; Rubin et al., 2014a).

Currently, neuroimaging features can be recorded and stored in a formalized format such as the DICOM (Digital Imaging and Communications in Medicine) SR (Structured Report) (Clunie, 2000; Clunie, 2007) and the Annotation and Imaging Markup (AIM) model (Channin et al., 2010): the DICOM SR formalizes the representation of radiological observations by introducing a set of rules that constrain concepts organization and a vocabulary (i.e. codes and associated code meanings) covering the domain of imaging observations. DICOM SR includes measurements and qualitative assessments, their relationships with image evidence and with the clinical interpretation of the clinician. The AIM model is an information model and an XML-based file format to describe the minimal information necessary to record image annotations. This information model has introduced the most relevant entities used in image annotation.

These standard formats enable the description of the content of medical images. Unfortunately, they are not suitable to support logic-based reasoning; these formats are based on coded terms and most of them have no semantic axioms to specify their meanings and the relationships between them. As a consequence, only searches based on keywords can be handled on decision making tools that exploit these standard formats. Rubin et al., Levy MA and many

Amdouni, E. and Gibaud, B.
Semantic Representation of Neuroimaging Observations: Proof of Concept based on the VASARI Terminology.
DOI: 10.5220/0006931100630074
In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 2: KEOD*, pages 63-74
ISBN: 978-989-758-330-8

others researchers have promoted the use of formal ontologies (Rubin et al., 2009; Levy et al., 2009; Kahn J.R. et al., 2011; Van Soest et al., 2014) for automatic reasoning on these data models and overcome the current limitations listed above.

This work is based on the assumption that the use of semantic web technologies, in particular ontologies as well as reasoning capabilities, can make the meaning of neuroimaging assessments more explicit and facilitate their advanced exploitation and interpretation. We are convinced that this approach can be beneficiary to the whole domain of neuroimaging, and beyond to the whole domain of biomedical imaging. However, this work focuses on a limited domain, namely the domain covered by the VASARI terminology (for Visually Accessible Rembrandt Images terminology) [1], used as a proof of a concept.

The VASARI terminology is a controlled vocabulary that describes thirty observations of high grade cerebral gliomas (glioblastoma multiform or GBM) in conventional Magnetic Resonance Imaging (MRI) images. Its main objective consists in standardizing brain tumors description and facilitating their interpretation by neuro-radiologists. The VASARI terminology was developed by experts in neuro-radiology who have considered the majority of possible assessments based on MRI. The validation of this corpus of imaging features was realized by eight experienced neuro-radiologists from distinct institutions.

Our work has three main contributions. First, making the meaning of VASARI features explicit via the design and the implementation of a specialized ontology, called VASARI ontology. Second, providing a semantic annotation tool that automatically translates VASARI features into instances of the VASARI ontology. This tool was applied to a VASARI corpus called REMBRANDT data set (Repository of Molecular Brain Neoplasia Data) (Madhavan et al., 2009). Third, performing semantic queries and reasoning tasks on these data to show how this semantic description of VASARI features can facilitate the interpretation of the content of medical images.

The remaining of this article is organized as follows: in Section Material and Methods, we describe the methodology for the design of the VASARI ontology. Section Results provides, first, an overview of the architecture and development details of both the developed ontology and the semantic annotation tool. Second, it demonstrates how we can automatically store and semantically manipulate RDF (Resource Description Framework) data of the REMBRANDT data set. In Section Discussion, we explain some modeling choices that we have made regarding ontology design and implementation, list accomplished work and enumerate remaining problems.

[1] https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project

2 MATERIAL AND METHODS

2.1 Design of the VASARI Ontology

The VASARI ontology was designed according to the realism-based approach proposed by Ceusters and Smith (Ceusters and Smith, 2005; Smith, 2006; Smith and Ceusters, 2010). Our modeling methodology is composed of five main steps that can be outlined as follows. First, we analyzed for each VASARI feature Fi the meaning of the studied aspect and sorted its possible configurations to establish the list of possible values allowed for each criterion. Second, we identified and described the key real entities that are involved in each criterion. Third, we related entities to existing ontologies, most of which coming from the OBO foundry and aligned onto the Basic Formal Ontology (Bittner and Smith, 2004; Smith et al., 2005b). When needed new ontology classes were specified. Fourth, we specified and defined the axioms characterizing these entities and relations between them. Finally, we made sure that all possible configurations for each feature Fi can be modeled in a formal way.

In our work, we described the thirty VASARI features but this paper focuses on seven of them, namely: lesion location, lesion side, enhancement quality, proportion nCET (non contrast enhanced tumor), cortical involvement, extent of resection of enhancing tumor and lesion size (see Table 1).

After a deep analysis of the meaning of the VASARI features and the identification of the different entities that they involve, we have proceeded with their formal description. This step was not a trivial task given that we faced several modeling problems summarized in Table 2 and more detailed in (Amdouni and Gibaud, 2016). Modeling problems concern: negative findings (MP1) (Ceusters et al., 2006), spatial knowledge (MP2) (Bennett et al., 2013) and complex entities representation (MP3).

We have used the version 2 of the Basic Formal Ontology (BFO) as a foundation for the VASARI ontology, thus facilitating the integration of specialized ontologies that come from the Open Biological and Biomedical Ontologies (OBO) foundry (Smith et al., 2007). In particular, we reused the following ontologies: the Foundational Model of Anatomy (FMA) (Smith et al., 2006), the Information Artifact Ontology (IAO) (Ceusters, 2012), the Phenotypic Quality Ontology (PATO) (Mungall et al., 2007), the Open

Table 1: Selected subset of the VASARI features defined in the VASARI terminology.

Feature	Feature definition	feature
F1.Lesion location	Location of lesion geographic epicenter; the largest component of the tumor either contrast enhancing or non contrast enhancing.	frontal, parietal, temporal, occipital, corpus callosum, thalamus
F2.Lesion side	Side of lesion epicenter.	right, central, bilateral
F4.Enhancement quality	Qualitative degree of contrast enhancement is defined as having all or portions of the tumor that demonstrate significantly higher signal on the post contrast T1W images compared to pre contrast T1W images.	n/a, none, mild, marked
F6.Proportion nCET	What proportion of the entire tumor is non enhancing? Non enhancing tumor is defined as regions of T2W hyper intensity that are associated with mass effect and architectural distortion, including blurring of the gray-white interface.	n/a, 0%, >5%, 6-33%, 34-67%, 68-95%, >95%, 100%, indeterminate.
F20.Cortical involvement	Non-enhancing o enhancing tumor extending to the cortical mantle, or cortex is no longer distinguishable relative to subjacent tumor.	no, yes
F26.Extent of resection of enhancing tumor	Using the first postoperative scan (contrast enhancing MR imaging) assessed for tumor residual estimating the proportion of enhancing tumor. Total resection component should be scored to 100%, subtotal resection of enhancing tissue should be scored accordingly.	n/a, 0%, >5%, 6-33%, 34-67%, 68-95%, >95%, 100%, indeterminate
F29.Lesion size	Largest perpendicular (x-y) cross section diameter of T2 signal abnormality (longest dimension X perpendicular) measured on single sectional image only.	unidimensional, largest diameter in centimeters

Biomedical Investigations (OBI) (Brinkman et al., 2010), the Ontology for General Medical Sciences (OGMS) (Scheuermann et al., 2009), the Unit Ontology (UO) (Gkoutos et al., 2012) and the Relation ontology (RO) (Smith et al., 2005a); in our work we have considered that RO relations are integrated under the BFO 2 ontology. Note that ontologies acronyms will be used in the rest of the paper.

2.2 Design of the Experimental Work

In our experimental work, we have developed a semantic annotation software of VASARI data using the VASARI ontology. This software enables the user to transform the informal description of the 30 VASARI features into a formal one. This software was applied to a corpus of VASARI data called the REMBRANDT repository. The resulting semantic data set was used to evaluate the added value of our work; especially, we performed some reasoning tasks by formulating semantic queries as well as some consistency tests to detect inconsistent assertions.

2.3 Presentation of the REMBRANDT Repository

The REMBRANDT repository is freely accessible on this link16. The ultimate objective of the REMBRANDT data set is to facilitate the discovery of significant correlations between clinical and genomic information in order to provide patients with more personalized treatments in the clinical context. The REMBRANDT data set contains 30 VASARI features labeled by 3 radiologists that concern 34 patients with GBM tumors. Features values are stored in an Excel

Table 2: Modeling problems.

Modeling problem	VASARI expressions	Use cases and examples
MP1: How to represent negative neuroimaging observations that indicate the non-existence of a dependent continuant (category C1) or an independent continuant (category C2): bfo:Quality or bfo:Disposition?	Negative findings are expressed by the use of negative qualifiers as for example none or expressions as indeterminate, not applicable, without etc.	Non-existence of an entity or a quality/disposition. Examples: C1:Johns cerebral tumor is without an enhancing region. C2:Johns cerebral tumor is not edematous. Or Johns cerebral tumor is not infiltrative.
MP2: How to ensure a faithful representation of the studied pathological structure and describe how its components are situated in space?	Containment is denoted by natural language expressions like within, portion of, comprise of whereas overlapping is denoted by the term invasion. The proximity of a given entity to another one is expressed with terms such as surrounding and adjacency, and separation is denoted by terms like not contiguous and separated.	Spatial location of existing entities, entities that are related to each other, entities that are separated from each other, entities that are adjacent to each other, etc. Examples: Johns cerebral tumor epicenter is in the parietal lobe. A cerebral tumor has part a cerebral tumor margin.
MP3: How to encode complex entities as for example derived measurements (proportions of volume measurements, length measurements, etc.), and associate them to their corresponding clinical findings?	Volume proportions: 0%, <5%, 6-33%, 34-67%, 68-95%, >95%, 100%. Or, two-dimensional length (x,y) that represents cross-sectional diameters; scores are between <0.5 cm and >8cm.	The representation of the extent of the resection of a given cerebral tumor component (nCET, necrotic, etc.). Examples: 35% of Johns cerebral tumor is enhanced, Johns cerebral tumor of size 1cm2cm.

document where each spreadsheet contains evaluations asserted by a radiologist.

2.4 Implementation Details

We have designed the VASARI ontology in the Ontology Web Language 2 (OWL2) format using the version 5 of the Prot*égé* tool (Tudorache et al., 2013). To extract modules from the relevant OBO ontologies we used the Ontofox web interface (Smith et al., 2007). Our semantic annotation tool was developed with the JAVA language and created using the NetBeans IDE 8.0.2 programming environment. To design the semantic annotation software, we used the version 2.8.3 of the JENA35 semantic web programming framework and the version 2.3.2 of the JENA-Pellet reasoner engine (external engine of the JENA API) (Sirin et al., 2007) to perform automatic reasoning tasks. To execute some SPARQL queries, we have used the CORESE tool 3.2 [2]. Note that instance data are serialized in the RDF/XML format.

3 RESULTS

3.1 The VASARI Ontology

The VASARI ontology imports eight ontology modules and contains around 570 OWL classes and 120 properties. Figure 1 shows that the four major semantic aspects that constitute the VASARI domain, namely: pathological structures, anatomical localization, qualities and dispositions, and measurements. In

[2] http://wimmics.inria.fr/corese

this section, classes are represented in *Italic* and relationships between them in **Bold**.

In our proposed ontology, a *vasari:CerebralTumor* **is_a** *vasari:CerebralPathologicalStructure*. Different regions of the cerebral tumor are introduced with the entity *vasari:CerebralTumorComponent* that is defined as a *vasari:CerebralPathologicalStructure* and **bfo:continuant_part_at_some_time** some *vasari:CerebralTumor*. In our semantic model and as it is defined in the VASARI terminology, we considered that a *vasari:CerebralTumor* can be composed of four basic components that characterize the brain tissue abnormality: *vasari:EnhancingCerebralTumorComponent*, *vasari:NonEnhancingCerebralTumorComponent*, *vasari:NecroticCerebralTumorComponent* and *vasari:CerebralEdemaComponent*.

The following paragraphs describe how the specific domains requirements mentioned in the Material and Methods Section were addressed in the VASARI ontology.

Modeling Problem 1: In order to respect the basic principles of realist ontology, no instance should be created when no concrete entity exist in reality. The solution that we proposed uses an equivalentclass axiom (condition if and only if involving some negative somevaluesfrom assertion). Thus, classes are defined as follows:

- *Case category C1*: A *vasari:CerebralTumorComponentNotLocatedInBrainCortex* [definition] **is_a** *vasari:CerebralTumorComponent* and not (**bfo:located_in_at_some_time** some *vasari:CerebralCortex*), an *vasari:EnhancingCerebralTumorWithoutNonEnhancingCerebralTumorComponent* [definition] **is_a** *vasari:EnhancingCerebralTumor* and not (**bfo:has_continuant_part_at_some_time** some *vasari:NonEnhancingCerebralTumorComponent*).
- *Case category C2*: A *vasari:NonCysticCerebralTumorComponent* [definition] **is_a** *vasari:CerebralTumorComponent* and not (**bfo:has_quality_at_some_time** some *pato:Cystic*), a *vasari:NonEnhancingCerebralTumorComponent* [definition] **is_a** *vasari:CerebralTumorComponent* and not (**bfo:has_disposition_at_some_time** some *vasari:DispositionToBeEnhancing*).

Modeling Problem 2: Three main spatial relations are modeled:

- *Containment relation*: We employed the spatial relation **bfo:located_in_at_some_time** and the foundational relation **bfo:part_of_continuant_at_some_time**.
We suppose that C1 and C2 are classes of continuants. As asserted in BFO, *C1* **bfo:part_of_continuant_at_some_time** *C2* means that for every particular c1, if c1 **instance_of** *C1* then there is some c2 such that c2 **instance_of** *C2* and c1 **bfo:part_of_continuant_at_some_time** c2, *C1* **bfo:located_in_at_some_time** *C2* asserts that for every c1 if c1 **instance_of** *C1*, then there is some c2 **instance_of** *C2* and c1 **bfo:located_in_at_some_time** c2. In our ontology, we used the relation **bfo:located_in_at_some_time**, for example to associate a particular *vasari:CerebralTumorEpicenter* to its specific *vasari:LobeOfCerebralHemisphere* and the relation **bfo:has_continuant_part_at_some_time** to define that the *vasari:EdematousCerebralTumor* [definition] **is_a** *vasari:CerebralTumor* and (**bfo:has_continuant_part_at_some_time** some *vasari:CerebralEdemaComponent*).
- *Overlapping vs adjacency relation*: We employed the spatial relation **ro:adjacent_to** to express that two continuants do not share a common spatial region and we defined the relation **vasari:overlaps** to represent the case of overlapping. As described in Table 1, F20 evaluates the location of the cerebral tumor regarding the cerebral cortex. To describe these different situations and ensure a correct classification of the cerebral tumor, we have defined the following classes: *vasari:CerebralTumorInvadingBrainCortex* [definition] *vasari:CerebralTumor* and (**vasari:overlaps** some *vasari:CerebralCortex*), *vasari:CerebralTumorAdjacentToBrainCortex* [definition] *vasari:CerebralTumor* and (**ro:adjacent_to** some *vasari:CerebralCortex*).
- *Separation relation*: We used the relational quality *vasari:ContiguousWithCerebralTumor* with the logical negation operator (not) to qualify and identify cerebral component that are separated from the cerebral tumor, as for example: *vasari:SatelliteLesion* [definition] *vasari:CerebralPathologicalStructure* and (not (**bfo:has_quality_at_some_time** some *vasari:ContiguousWithCerebralTumor*)) and (**bfo:has_disposition_at_some_time** some *vasari:DispositionToBeEnhancing*).
- *Separation relation*: We used the relational quality *vasari:ContiguousWithCerebralTumor* with the logical negation operator (not) to

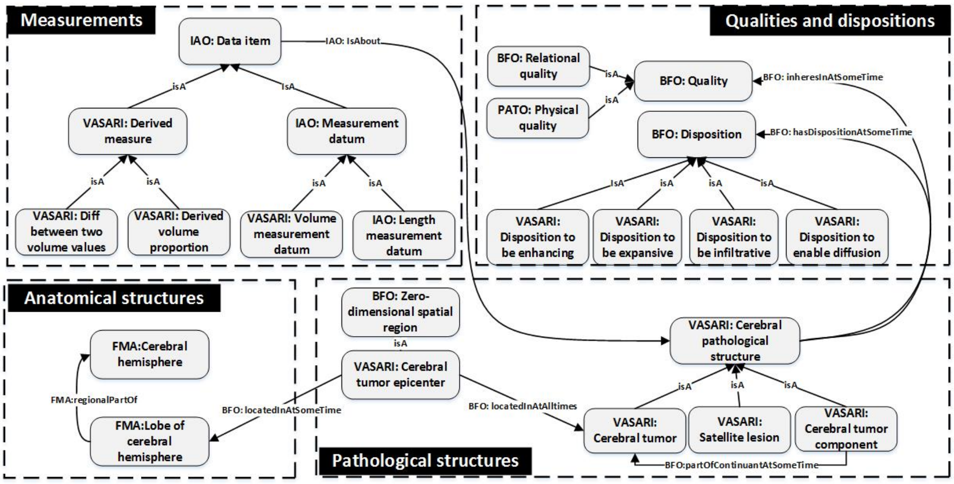


Figure 1: The basic pattern of the main classes in the VASARI ontology.

qualify and identify cerebral component that are separated from the cerebral tumor, as for example: *vasari:SatelliteLesion* [definition] *vasari:CerebralPathologicalStructure* and (not (**bfo:has_quality_at_some_time** some *vasari:ContiguousWithCerebralTumor*)) and (**bfo:has_disposition_at_some_time** some *vasari:DispositionToBeEnhancing*).

Modeling Problem 3: The representation of the extent of the resection of an enhancing cerebral tumor component appears to be simple data to represent, but in reality it involves hidden information that are not explicit in the definition of the feature. To model this kind of feature, we consider that the *vasari:EnhancingCerebralTumorComponent* will not preserve its identity after and before the surgery. Thus, we identified two distinct entities: *vasari:EnhancingCerebralTumorComponentBeforeSurgery* [definition] **is_a** *vasari:EnhancingCerebralTumorComponent* and (**bfo:part_of_continuant_at_some_time** some *vasari:EnhancingCerebralTumorComponent*) and (**bfo:is_specified_input_of** some *vasari:ResectionOfEnhancingCerebralTumorComponent*). *vasari:EnhancingCerebralTumorComponentAfterSurgery* [definition] **is_a** *vasari:EnhancingCerebralTumorComponent* and (**bfo:part_of_continuant_at_some_time** some *vasari:EnhancingCerebralTumorComponentBeforeSurgery*) and (**bfo:is_specified_output_of** some *vasari:ResectionOfEnhancingCerebralTumorComponent*).

We considered that the measured quality, i.e. the volume, is the same, but the measured volume values are distinct: *vasari:volumeMeasurementDatumOfEnhancingCerebralTumorAfterSurgery* [definition] *vasari:VolumeMeasurementDatum* and (**bfo:is_about** some *vasari:EnhancingCerebralTumorComponentAfterSurgery*), *vasari:volumeMeasurementDatumOfEnhancingCerebralTumorBeforeSurgery* [definition] *vasari:volumeMeasurementDatum* and (**bfo:is_about** some *vasari:enhancingCerebralTumorComponentBeforeSurgery*). The interpretation of extreme values of F26 will be as follows:

- %0 means that the enhancing tumor component is totally preserved and that the measured volume value before the surgery is the measured volume value after the surgery.
- %100 means that the enhancing cerebral tumor component is totally resected, thus the cerebral tumor component is classified as a *vasari:NonEnhancingCerebralTumorComponentAfterSurgery*.

3.2 Semantic Annotation Software of VASARI Data

The annotation software begins by reading as an input the set of imaging features values of the REMBRANDT repository and the VASARI ontology schema. Then, to semantically annotate the data the software realizes four main tasks. First, it instantiates the VASARI ontology based on the VASARI labeled values. Second, it describes imaging features by creating RDF triples that establish semantic links between instances. Third, it adds these triples as statements in an RDF graph. Fourth, it serializes data in the RDF/XML grammar and records the RDF graph in memory or in a JENA triple database (TDB). Note that the software stores separately the schema and the instance data (denoted by the terms Tbox and Abox in the following paragraph). It returns the generated RDF graph of the whole REMBRANDT dataset in 1.06 s (0.47s per radiologist).

3.3 Semantic Exploitation of the VASARI Annotation Data

The developed software bases its reasoning on an inferred model generated with a reasoner. The information stored in the inferred graph is contained in a knowledge base (KB) that can be exploited in two ways: 1) accessed via SPARQL queries to retrieve data based on their semantics 2) checked via consistency tests.

Querying a knowledge base: Let us consider this example to illustrate inference capabilities and demonstrate how the semantic format enables the exploitation of the anatomical knowledge coming from the FMA ontology. We suppose that a given KB is composed of an Abox that contains the following assertions: (A1) cte instance_of CerebralTumorEpicenter, (A2) tl instance_of TemporalLobe, (A3) rch instance_of RightCerebralHemisphere, (A4) cte bfo:located_in_at_some_time tl, (A5) tl regional_part_of rch. A Tbox that contains: (T1) RightTemporalLobe [equivalentClass] TemporalLobe and regional_part_of some RightCerebralHemisphere, (T2) regional_part_of [equivalentProperty] part_of_continuant_at_some_time.

Based on the assertions of the Abox and on the (T1) axiom we can deduce that: (A6) tl instance_of RightTemporalLobe (see Figure 3). Using (T2), (T3) and (A6), we can infer: (A7) cte bfo:located_in_at_some_time rtl (see figure 2).

Figure 3 presents an example of SPARQL query that retrieves the location of cerebral tumors (?lobe, ?cerebral_hemisphere) and their corresponding length measurements (?long_axis_value, ?perpendicular_long_axis_value). The lower part of the figure presents the results returned by the CORESE semantic query tool.

Validation of a knowledge base: The pellet JENA reasoner enables the detection of conflicts in the knowledge content; thus we have exploited this capacity to perform a global check across the KB and looked for inconsistencies between the radiologists assertions. For example lets suppose that (A1) ct instance_of CerebralTumor and that one radiologist said that (A2) ct is_a HemorrhagicCerebralTumor however, another radiologist said that (A3) ct is_a NonHemorrhagicCerebralTumor. Reasoning on this knowledge base is impossible given because the two classes HemorrhagicCerebralTumor and NonHemorrhagicCerebralTumor are defined as disjoint classes in the Tbox; this means that they cannot share the same set of instances (i.e., ct). As a consequence, a classification error should occur when the reasoner performs inference tasks.

The result of a consistency checking is provided via the object ValidityReport of the JENA API. This data structure encapsulates all detected inconsistent axioms and assertions. To generate explanations about inconsistencies, we have used the method explainconsistency(). The output generated after the execution of this method consists on the listing of the set of all involved axioms; Figure 4 depicts an example of inconsistency that is caused by the assignment of two different length measures to the same cerebral tumor. This multiple attribution violates the owl:functionalProperty axiom of the property iao:has_measurement_value that is declared as functional.

4 DISCUSSION

In our work, we followed the realism-based approach to describe the neuroimaging reality on the side of the patient which appeared to us the most relevant methodology in the context of the biomedical research; the realism-based approach is being adopted by a growing community of researchers in the medical context. Actually, medical terminologies such as the DICOM SR and the AIM model do not refer to concrete existing phenomena on the side of the patient, but they only code medical statements in a formal way. The adoption of the realism-based approach enabled us to provide a faithful representation of imaging features by considering both the universals level (e.g. cerebral tumor concept) and instances of universals level (e.g. Davids cerebral tumor). To follow this modeling perspective, we have used two foundational and realism-based ontologies; namely the BFO ontology to describe existing entities and relationships between them. The use of BFO ontology has facilitated the integration of heterogeneous knowledge from different ontologies that are specialized in anatomy, quality phenotypes, measurements, etc.

The developed ontology answers to some challenging points that are highlighted in many papers (Ceusters et al., 2006; Cimino, 2006; Levy et al., 2012) mainly: 1) the formalization of the description of neuroimaging information via the use of specialized ontologies; we can cite the example of the FMA ontology that allowed the description of some clinical statements or the BFO and RO ontologies that helped us in

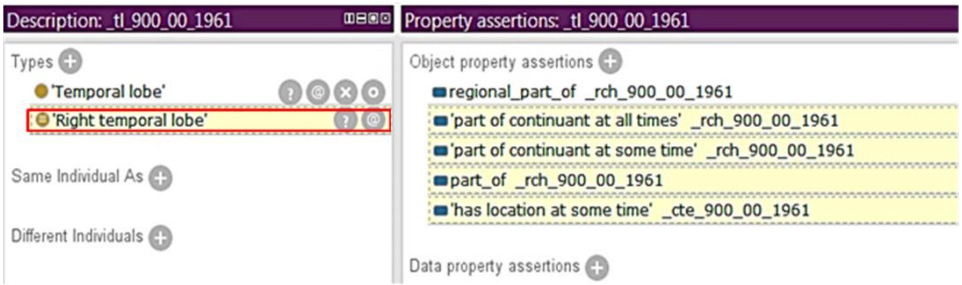
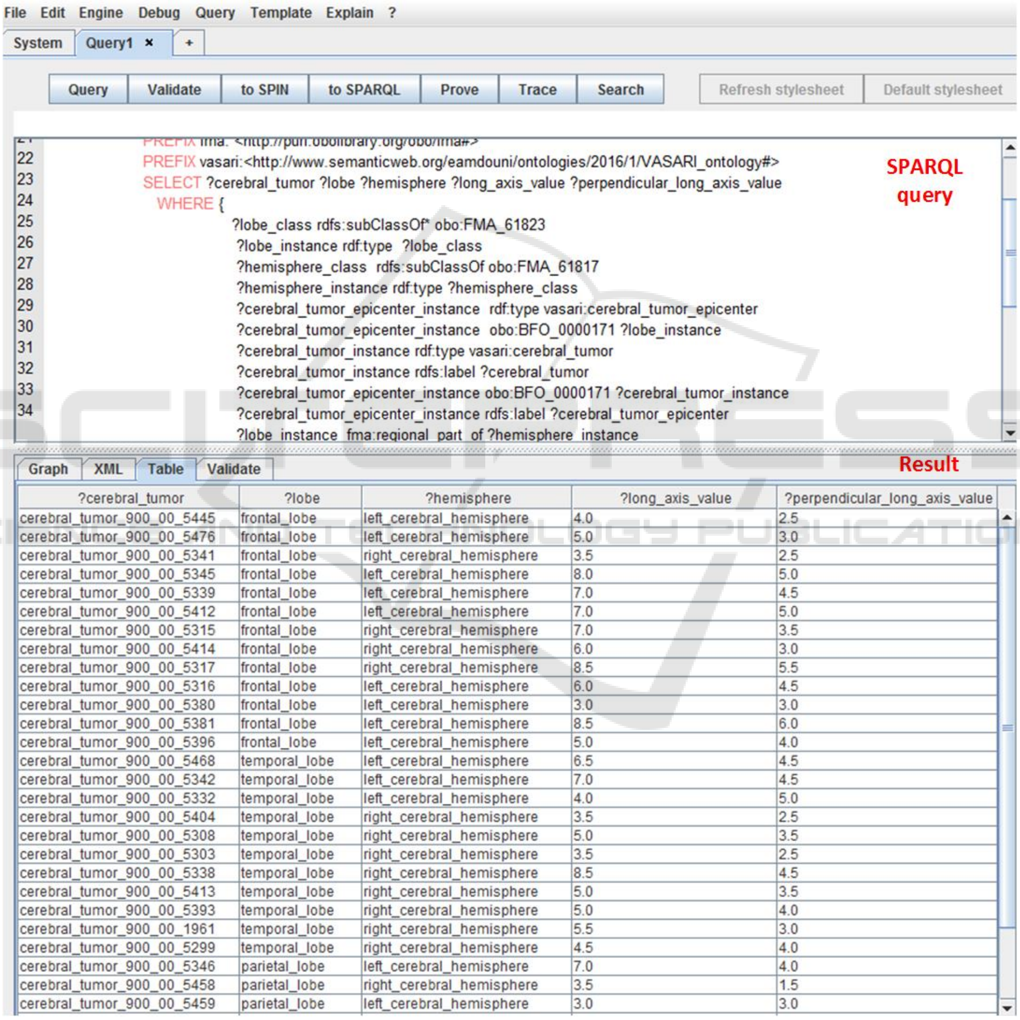


Figure 2: Illustrative example of semantic exploitation of anatomical knowledge from the FMA ontology (protg tool); instances names are automatically generated with our developed annotation tool, the ID of the patient 900_00_1961 is included to refer to existing entities on the side of the patient.



?cerebral_tumor	?lobe	?hemisphere	?long_axis_value	?perpendicular_long_axis_value
cerebral_tumor_900_00_5445	frontal_lobe	left_cerebral_hemisphere	4.0	2.5
cerebral_tumor_900_00_5476	frontal_lobe	left_cerebral_hemisphere	5.0	3.0
cerebral_tumor_900_00_5341	frontal_lobe	right_cerebral_hemisphere	3.5	2.5
cerebral_tumor_900_00_5345	frontal_lobe	left_cerebral_hemisphere	8.0	5.0
cerebral_tumor_900_00_5339	frontal_lobe	left_cerebral_hemisphere	7.0	4.5
cerebral_tumor_900_00_5412	frontal_lobe	left_cerebral_hemisphere	7.0	5.0
cerebral_tumor_900_00_5315	frontal_lobe	right_cerebral_hemisphere	7.0	3.5
cerebral_tumor_900_00_5414	frontal_lobe	right_cerebral_hemisphere	6.0	3.0
cerebral_tumor_900_00_5317	frontal_lobe	right_cerebral_hemisphere	8.5	5.5
cerebral_tumor_900_00_5316	frontal_lobe	left_cerebral_hemisphere	6.0	4.5
cerebral_tumor_900_00_5380	frontal_lobe	left_cerebral_hemisphere	3.0	3.0
cerebral_tumor_900_00_5381	frontal_lobe	left_cerebral_hemisphere	8.5	6.0
cerebral_tumor_900_00_5396	frontal_lobe	left_cerebral_hemisphere	5.0	4.0
cerebral_tumor_900_00_5468	temporal_lobe	left_cerebral_hemisphere	6.5	4.5
cerebral_tumor_900_00_5342	temporal_lobe	left_cerebral_hemisphere	7.0	4.5
cerebral_tumor_900_00_5332	temporal_lobe	left_cerebral_hemisphere	4.0	5.0
cerebral_tumor_900_00_5404	temporal_lobe	right_cerebral_hemisphere	3.5	2.5
cerebral_tumor_900_00_5308	temporal_lobe	right_cerebral_hemisphere	5.0	3.5
cerebral_tumor_900_00_5303	temporal_lobe	right_cerebral_hemisphere	3.5	2.5
cerebral_tumor_900_00_5338	temporal_lobe	right_cerebral_hemisphere	8.5	4.5
cerebral_tumor_900_00_5413	temporal_lobe	right_cerebral_hemisphere	5.0	3.5
cerebral_tumor_900_00_5393	temporal_lobe	right_cerebral_hemisphere	5.0	4.0
cerebral_tumor_900_00_1961	temporal_lobe	right_cerebral_hemisphere	5.5	3.0
cerebral_tumor_900_00_5299	temporal_lobe	right_cerebral_hemisphere	4.5	4.0
cerebral_tumor_900_00_5346	parietal_lobe	left_cerebral_hemisphere	7.0	4.0
cerebral_tumor_900_00_5458	parietal_lobe	right_cerebral_hemisphere	3.5	1.5
cerebral_tumor_900_00_5459	parietal_lobe	left_cerebral_hemisphere	3.0	3.0

Figure 3: Execution result to a SPARQL query on observations of the radiologist1 in the REMBRANDT repository (demonstration with the CORESE tool). FMA_61823 denotes the LobeOfCerebralHemisphere, FMA_61817 denotes the CerebralHemisphere and BFO_0000171 denotes the relation bfo:located_in_at_some_time.

the description of foundational (i.e., is a, part of) and spatial relations (i.e., location, adjacent to) between pathological structures, 2) the representation of negative neuroimaging observations via the use of OWL axioms and 3) the representation of complex entities was made more explicit by referring to the involved concrete entities; but we have faced some difficulties to represent mathematical expressions with the IAO

Table 3: Alignment of ontologies to represent VASARI features.

VASARI feature	Main involved classes	Main involved relations
F1:lesion location	vasari:CerebralTumorEpicenter, fma:LobeOfCerebralHemisphere, fma:Brainstem, fma:Cerebellum	bfo:located_in_at_some_time fma:regional_part_of
F2:lesion side	vasari:CerebralTumorEpicenter, fma:CerebralHemisphere, fma:MedianSagittalPlane	bfo:located_in_at_some_time
F4:enhancement quality	pato:Volume, vasari:CerebralTumorComponent, vasari:VolumeMeasurementDatum, obi:ValueSpecification	iao:is_quality_measurement_of, bfo:is_specified_input_of, bfo:is_specified_output_of, bfo:is_about
F5: proportion nCET	pato:Volume, vasari:CerebralTumorComponent, vasari:VolumeMeasurementDatum, obi:ValueSpecification	iao:is_quality_measurement_of, bfo:is_specified_input_of, bfo:is_specified_output_of, bfo:is_about
F20:cortical involvement	vasari:CerebralTumor, fma:CerebralCortex	vasari:overlaps ro:adjacent_to
F26:extent of resection of enhancing tumor	pato:Volume, vasari:VolumeMeasurementDatum, uo:VolumeUnit, vasari:RatioValueSpecification, vasari:ProportionOfEnhancingAndRemovedTumor, vasari:EnhancingCerebralTumor-Component, vasari:ResectionOfEnhancingCerebralTumorComponent	iao:is_quality_measurement_of, bfo:has_measurement_unit_label, iao:has_measurement_value, bfo:is_about, obi:is_specified_input_of, obi:is_specified_output_of, vasari:has_specified_denominator_-value, vasari:has_specified_nominator_-value, ro:derives_from
F29 and F30: lesion size	obi:LengthMeasurementDatum, uo:LengthUnit, pato:Quality	iao:is_quality_meausrement_of, iao:has_measurement_value, iao:has_measurement_unit_label, iao:is_about

ontology, thus we think that it will be interesting to extend it to cover this kind of information that is needed in the definition of biomedical experiments.

It is important to note that our approach is not limited to the translation of VASARI features resulting from the subjective assessment of MRI images by human neuro-radiologists. In contrast, it would express its full value if it were implemented as a complement to an automated or semi-automated image analysis system (Velazquez et al., 2015). For example systems that are described in these papers (Rubin et al., 2014b; Porz et al., 2014) enable to segment the various parts of the tumor, and to automatically determine their anatomical environment (e.g. what anatomical structures they are contained in, or they overlap or they are adjacent to). Such mereotopological properties could be directly translated in semantic form using the relationship discussed above. Similarly, the volume measurements and the derived proportions could be generated automatically and with a better accuracy than through the current subjective assessments. VASARI features could be derived from the detailed image-based observations and measurements rather precede them, and our model provides the conceptual basis to make such enrichment of image processing systems.

The Radiology Reading Room of the Future (Gillies et al., 2015) will entail a reading room where in practicing radiologists interact with picture archiving and communication system software to identify, segment, and extract features from regions of interest. If prior studies obtained in the same patient are available, the previous regions of interest will be automatically identified by the reading software. As part of the reading, the extracted size, shape, location, and textural features will be automatically uploaded to a shared database and algorithmically compared with prior images to enable more precise diagnoses.

The first limit of our proposal is that it is implemented in OWL and thus it does not generate temporalized instances (Smith et al., 2006). We think that

```
The model is not consistent and some conflicts are detected:
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:OBO_NS="http://purl.obolibrary.org/obo/"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:RO_NS="http://www.ifomis.org/obo/ro/"
    xmlns:VASARI_NS="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#">
  <owl:FunctionalProperty rdf:about="http://purl.obolibrary.org/obo/IAO_0000004"/>
  <rdf:Description rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_paldmd_900_00_1961">
    <OBO_NS:IAO_0000004 rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
    >5.5</OBO_NS:IAO_0000004>
    <OBO_NS:IAO_0000004 rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
    >5.0</OBO_NS:IAO_0000004>
  </rdf:Description>
</rdf:RDF>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:OBO_NS="http://purl.obolibrary.org/obo/"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:RO_NS="http://www.ifomis.org/obo/ro/"
    xmlns:VASARI_NS="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#" >
  <rdf:Description rdf:about="http://purl.obolibrary.org/obo/IAO_0000004">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_paldmd_900_00_1961">
    <OBO_NS:IAO_0000004 rdf:datatype="http://www.w3.org/2001/XMLSchema#float">5.5</OBO_NS:IAO_0000004>
    <OBO_NS:IAO_0000004 rdf:datatype="http://www.w3.org/2001/XMLSchema#float">5.0</OBO_NS:IAO_0000004>
  </rdf:Description>
</rdf:RDF>
```


Figure 4: Example of a validation output: detection of a logical contradiction, IAO_000004 denotes the relation iao:has_measurement_value and _paldmd_900_00_1961 refers to the measurement value of the perpendicular longest axis of the patient 900_00_1961.

taking into consideration the temporal aspect in the representation of neuroimaging features is needed especially in longitudinal imaging studies to, for example, evaluate cancer treatment response. In this context, we recommend to select a logic-based language that is capable to represent n-ary relationships. The second limit is intrinsic to the problem of logical contradiction that is due to the fact that radiologists describe what they observe based on their thoughts and experiences. As a consequence they may describe differently the reality and produce different clinical records about the identified entities. To resolve the disagreement in interpretation many medical systems require preserving in data entry a single correct value for the evaluated feature to facilitate the aggregation of data. According to Rector et al., electronic medical records should allow the presence of conflicting statements, multiple measurements, etc. to faithfully reflect the reality of clinical practice. Thus, they propose three types of foundational information models to describe (a) the medical records, (b) the state of the patient and (c) clinical care (Rector et al., 1991) and they consider that a meta-language should be used to separate between what can be said and what actually occurs, and avoid the problem of inconsistency. Smith et al. have mentioned that even the adoption of a meta-language cannot remove errors because medical dialogues are also subject to error (Smith et al., 2006). In our work, we have stored (a) observations in separate data sets and we have not included (b and c) the meta-observation level.

The experimental work regarding the VASARI ontology shows that the semantic representation of neuroimaging features can enhance search operations; e.g. the exploitation of the VASARI ontology with the DL reasoner can resort the list of cerebral tumor epicenters that are located in the right temporal lobe by retrieving cerebral tumor epicenters that are located in the temporal lobe and the right hemisphere. Certainly, such classification task cannot be obtained with non semantic representations such as DICOM SR and the AIM model. We believe that our work can be reused in other image-based reasoning contexts as for example, RECIST criteria that base the tumor classification task (i.e. measurable and non-measurable lesions) on the knowledge of the location of the lesion and the calculation of its length. Added to the reasoning task, the consistency checking functionality offered by OWL reasoners can detect inconsistent statements that can be caused by inappropriate or erroneous diagnosis or treatment, in the clinical context, until now DICOM SR and the AIM model do not offer this semantic capability.

5 CONCLUSION

We believe that neuroimaging data should be held in a structured format that makes their meanings explicit to the systems and thus facilitate their comprehension as well as management. Semantic data about imaging features (measurement values, qualities, lesion components, lesion localization, etc.) are important (1) to support the clinical research on the development of new imaging biomarkers by combining clinical data with information coming from different medical domains (2) to improve the quality of the clinical healthcare that tend to provide personalized treatments to patients via the use of clinical guidelines that are based on evaluation criteria. In this paper, we employed the VASARI terminology as a proof of concept for the demonstration of the feasibility and the importance of making RDF and OWL data available to describe cerebral tumors observations and determining the key concepts and relationships that are central in their evaluation. Our work can be easily expanded to answer to other use cases; thanks to the modular aspect of the ontology and to the OWL language that is self-descriptive (concepts are textually and formally described in the ontology to guide users) and extendable.

REFERENCES

Amdouni, E. and Gibaud, B. (2016). Concept-based versus realism-based approach to represent neuroimaging observations. In *Keod: Proceedings of The 8th International Joint Conference On Knowledge Discovery, Knowledge Engineering and Knowledge Management-Vol. 2*, pages 179–185.

Bennett, B., Chaudhri, V., and Dinesh, N. (2013). A vocabulary of topological and containment relations for a practical biological ontology. In *International Conference on Spatial Information Theory*, pages 418–437. Springer.

Bittner, T. and Smith, B. (2004). Normalizing medical ontologies using basic formal ontology.

Brinkman, R., Courtot, M., Derom, D., Fostel, J., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., et al. (2010). Modeling biomedical experimental processes with OBI. *J. Biomedical Semantics*, 1(S-1):S7.

Ceusters, W. (2012). An information artifact ontology perspective on data collections and associated representational artifacts. In *MIE*, pages 68–72.

Ceusters, W., Elkin, P., and Smith, B. (2006). Referent tracking: The problem of negative findings. *Studies in health technology and informatics*, 124:741.

Ceusters, W. and Smith, B. (2005). Tracking referents in electronic health records. *Studies in health technology and informatics*, 116:71.

Channin, D., Mongkolwat, P., Kleper, V., Sepukar, K., and Rubin, D. (2010). The caBIG annotation and image markup project. *Journal of digital imaging*, 23(2):217–225.

Cimino, J. (2006). In defense of the desiderata. *Journal of biomedical informatics*, 39(3):299–306.

Clunie, D. (2000). *DICOM structured reporting*. PixelMed Publishing.

Clunie, D. (2007). DICOM structured reporting and cancer clinical trials results. *Cancer informatics*, 4:33.

Eisenhauer, E., Therasse, P., B. J., Schwartz, L., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., M., M., et al. (2009). New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer*, 45(2):228–247.

ESR (2011). Medical imaging in personalised medicine: a white paper of the research committee of the european society of radiology (esr). *Insights into imaging*, 2(6):621–630.

Gillies, R., Kinahan, P., and Hricak, H. (2015). Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577.

Gkoutos, G., Schofield, P., and Hoehndorf, R. (2012). The units ontology: a tool for integrating units of measurement in science. *Database*, 2012:bas033.

Grams, A., Gempt, J., Ringel, F., Soehngen, E., Astner, S., Schlegel, J., Meyer, B., Zimmer, C., and Forschler, A. (2014). Multimodal imaging to delineate tumor heterogeneity in cerebral gliomas. *Open Journal of Radiology*, 4(02):182.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.

Gutman, D., Cooper, L., Hwang, S., Holder, C., Gao, J., Aurora, T., Dunn Jr, W., Scarpace, L., Mikkelsen, T., Jain, R., et al. (2013). Mr imaging predictors of molecular profile and survival: multi-institutional study of the tcga glioblastoma data set. *Radiology*, 267(2):560–569.

Kahn J.R., C., Langlotz, C., Channin, D., and Rubin, D. (2011). Informatics in radiology: an information model of the dicom standard. *Radiographics*, 31(1):295–304.

Levy, M., Freymann, J., Kirby, J., Fedorov, A., Fennessy, F., Eschrich, S., Berglund, A., Fenstermacher, D., Tan, Y., G. X., et al. (2012). Informatics methods to enable sharing of quantitative imaging research data. *Magnetic resonance imaging*, 30(9):1249–1256.

Levy, M., OConnor, M., and Rubin, D. (2009). Semantic reasoning with image annotations for tumor assessment. In *AMIA Annual Symposium Proceedings*, volume 2009, page 359. American Medical Informatics Association.

Madhavan, S., Zenklusen, J., Kotliarov, Y., Sahni, H., Fine, H., and Buetow, K. (2009). Rembrandt: helping personalized medicine become a reality through integrative translational research. *Molecular cancer research*, 7(2):157–167.

Möller, M., Regel, S., and Sintek, M. (2009). Radsem: Semantic annotation and retrieval for medical images. In *European Semantic Web Conference*, pages 21–35. Springer.

Mungall, C., Gkoutos, G., Washington, N., and Lewis, S. (2007). Representing phenotypes in OWL. In *Proceedings of the OWLED 2007 Workshop on OWL: Experience and Directions*. Innsbruck, Austria Edited by Golbreich C, Kalyanpur A, Parsia B.

Oberkampf, H., Zillner, S., Bauer, B., and Hammon, M. (2012). Interpreting patient data using medical background knowledge. *ICBO*, 897:3.

Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R., Slotboom, J., Reyes, M., and Wiest, R. (2014). Multi-modal glioblastoma segmentation: man versus machine. *PloS one*, 9(5):e96873.

Rector, A.L., N. W. K. S. et al. (1991). Foundations for an electronic medical record. *Methods Inf Med*, 30(3):179–186.

Rubin, D., Mongkolwat, P., and Channin, D. (2009). A semantic image annotation model to enable integrative translational research. *Summit on translational bioinformatics*, 2009:106.

Rubin, D., Willrett, D., O'Connor, M., Hage, C., Kurtz, C., and Moreira, D. (2014a). Automated tracking of quantitative assessments of tumor burden in clinical trials. *Translational oncology*, 7(1):23–35.

Rubin, D., Willrett, D., O'Connor, M., Hage, C., Kurtz, C., and Moreira, D. (2014b). Automated tracking of quantitative assessments of tumor burden in clinical trials. *Translational oncology*, 7(1):23.

Scheuermann, R., Ceusters, W., and Smith, B. (2009). Toward an ontological treatment of disease and diagnosis. *Summit on translational bioinformatics*, 2009:116.

Seifert, S., Kelm, M., Moeller, M., Mukherjee, S., Cavallaro, A., Huber, M., and Comaniciu, D. (2010). Semantic annotation of medical images. In *SPIE medical imaging*, pages 762808–762808. International Society for Optics and Photonics.

Sirin, E., Parsia, B., Grau, B., Kalyanpur, A., and Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, 5(2):51–53.

Smith, B. (2006). From concepts to clinical reality: an essay on the benchmarking of biomedical terminologies. *Journal of biomedical informatics*, 39(3):288–298.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L., Eilbeck, K., Ireland, A., Mungall, C., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255.

Smith, B. and Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied ontology*, 5(3-4):139–188.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A., and Rosse, C. (2005a). Relations in biomedical ontologies. *Genome biology*, 6(5):R46.

Smith, B., Kumar, A., and Bittner, T. (2005b). Basic formal ontology for bioinformatics. *Journal of Information Systems*, pages 1–16.

Smith, B., Kusnierczyk, W., Schober, D., and Ceusters, W. (2006). Towards a reference terminology for ontology research and development in the biomedical domain. In *KR-MED*, volume 2006, pages 57–66.

Tudorache, T., Nyulas, C., Noy, N., and Musen, M. (2013). Webprotégé: A collaborative ontology editor and knowledge acquisition tool for the web. *Semantic web*, 4(1):89–99.

Van Soest, J., Lustberg, T., Grittner, D., Marshall, M., Persoon, L., Nijsten, B., Feltens, P., and Dekker, A. (2014). Towards a semantic pacs: using semantic web technology to represent imaging data. *Studies in health technology and informatics*, 205:166.

Velazquez, E., Meier, R., Dunn Jr, W., Alexander, B., Wiest, R., Bauer, S., Gutman, D., Reyes, M., and Aerts, H. (2015). Fully automatic gbm segmentation in the tcga-gbm dataset: Prognosis and correlation with vasari features. *Scientific reports*, 5.