



**HAL**  
open science

# Représentation sémantique des biomarqueurs d'imagerie dans le domaine médical

Emna Amdouni

► **To cite this version:**

Emna Amdouni. Représentation sémantique des biomarqueurs d'imagerie dans le domaine médical. Traitement des images [eess.IV]. Université de Rennes, 2017. Français. NNT : 2017REN1S124 . tel-01801778

**HAL Id: tel-01801778**

**<https://theses.hal.science/tel-01801778>**

Submitted on 28 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Bretagne Loire*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**  
*Mention : Informatique*

École doctorale Mathstic  
présentée par

**Emna AMDOUNI**

Préparée à l'unité de recherche IRT B<>com et l'équipe MediCIS, LTSI-  
INSERM 1099  
Institut de Recherche Technologique B<>COM  
UFR Informatique et Électronique (ISTIC)

---

**Représentation  
Sémantique des  
Biomarqueurs  
d'Imagerie en  
Imagerie Médicale**

**Thèse soutenue à Rennes  
le 07 décembre 2017**

devant le jury composé de :

**Mme Sylvie DESPRES**

Professeure à l'université de Paris 13 Nord/  
*rapporteur*

**Mr Djamal BENSLIMANE**

Professeur à l'université de Claude Bernard Lyon 1/  
*rapporteur*

**Mr Jean-Yves GAUVRIT**

Professeur à l'université de Rennes 1/ *examinateur*

**Mme Rim FAIZ**

Professeure à l'Institut des Hautes Etudes  
Commerciales de Carthage/ *examinatrice*

**Mr Bernard GIBAUD**

Chargé de recherches INSERM/ *directeur de thèse*



*"If you can not measure it, you can not improve it"*  
*Lord Kelvin (1883)*



# Abstract

In personalized medicine, radiological measurements and observations 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). Many radiology-pathology correlation studies show that quantitative and qualitative imaging features are associated to genetic alterations and gene expression. Therefore, suitable management of imaging phenotypes is needed to facilitate their use and reuse in multiple studies regarding radiological measurements. In literature, radiological measurements that characterize biological processes of imaged subjects are called imaging biomarkers. The main objective of this thesis is to propose an ontological conceptualisation of imaging biomarkers to make their meaning explicit and formal, and improve structured reporting of images. The first part of the thesis presents a generic ontology that defines basic aspects of the imaging biomarker concept, namely ; measured biological characteristic, measurement protocols and role in decision making application. The second part of the thesis addresses important semantic modeling challenges related to the description of neuro-imaging data using existing biomedical knowledge, as well as it proposes some “relevant” solutions to the most typical situations that need to be modeled in glioblastoma.

**Keywords :** knowledge representation, domain ontology, biomedical ontologies, imaging biomarker.



# Résumé

En médecine personnalisée, les mesures et les descriptions radiologiques jouent un rôle important ; en particulier, elles facilitent aux cliniciens l'établissement du diagnostic, la prise de décision thérapeutique ainsi que le suivi de la réponse au traitement. On peut citer à titre d'exemple, les critères d'évaluation RECIST (en anglais Response Evaluation Criteria in Solid Tumors). De nombreuses études de corrélation en radiologie-pathologie montrent que les caractéristiques d'imagerie quantitative et qualitative sont associées aux altérations génétiques et à l'expression des gènes. Par conséquent, une gestion appropriée des phénotypes d'imagerie est nécessaire pour faciliter leur utilisation et leur réutilisation dans de multiples études concernant les mesures radiologiques. Dans la littérature, les mesures radiologiques qui caractérisent les processus biologiques des sujets imagés sont appelées biomarqueurs d'imagerie. L'objectif principal de cette thèse est de proposer une conceptualisation ontologique des biomarqueurs d'imagerie pour rendre leur sens explicite et formel, et améliorer leur interprétation sous la forme de comptes rendus structurés. La première partie de la thèse présente une ontologie générique qui définit les aspects fondamentaux du concept de biomarqueur d'imagerie. La deuxième partie de la thèse traite des problèmes de modélisation sémantique liés à la description des données d'observation en neuro-imagerie en utilisant les connaissances biomédicales existantes. Ainsi, elle propose des solutions « pertinentes » aux situations les plus typiques qui doivent être modélisées dans le glioblastome.

**Mots-clés :** représentation des connaissances, ontologie de domaine, ontologies biomédicales, biomarqueur d'imagerie.



# Table des matières

<b>1</b>	<b>Introduction Générale</b>	<b>13</b>
<b>2</b>	<b>Contexte général du travail de la thèse</b>	<b>17</b>
2.1	Introduction	18
2.2	L'avènement de la médecine de précision	18
2.2.1	Les biomarqueurs, la stratification et les traitements appropriés au cœur de la médecine personnalisée	19
2.2.2	Promesses versus réalités de la médecine personnalisée	20
2.3	L'imagerie quantitative au service de la médecine de précision	20
2.3.1	Tomodensitométrie	21
2.3.2	Imagerie par Résonance Magnétique	23
2.3.3	Médecine nucléaire	29
2.4	L'ère des biomarqueurs d'imagerie	31
2.4.1	Définition du concept de biomarqueur d'imagerie	31
2.4.2	Utilité des biomarqueurs d'imagerie	32
2.4.3	Standardisation des biomarqueurs d'imagerie	39
2.4.4	Processus de développement des biomarqueurs d'imagerie	41
2.4.5	Classification des biomarqueurs d'imagerie	45
2.5	Besoin d'intégration des données d'imagerie pour supporter le développement des bio-banques d'images	47
2.6	Contexte d'étude : représentation des observations d'imagerie médicale	50
2.6.1	Description des besoins médicaux dans les applications dédiées à la délivrance de soins	50
2.6.2	Description des besoins médicaux dans les applications dédiées à la recherche	54
<b>3</b>	<b>État de l'art de la thèse</b>	<b>59</b>
3.1	Introduction	59
3.2	Représentation de la sémantique des données d'imagerie avec des standards informatiques	60
3.2.1	Comptes rendus structurés avec le standard DICOM	60
3.2.2	Sémantique reposant sur le modèle d'information AIM	67
3.3	Structures de données dans les systèmes à base de connaissances	70

3.3.1	Terminologies, thésaurus et ontologies . . . . .	70
3.3.2	Quelques standards fondamentaux du Web sémantique . . . . .	72
3.4	Représentation des connaissances biomédicales . . . . .	74
3.4.1	Format UMLS . . . . .	74
3.4.2	Librairie OBO . . . . .	75
3.5	Représentation de la sémantique des données d'imagerie avec des ontologies . . . . .	76
3.5.1	Le modèle Open Annotation . . . . .	76
3.5.2	L'ontologie OntoNeuroLOG . . . . .	78
3.5.3	L'ontologie QIBO . . . . .	79
3.5.4	Le Système BiomRKRS . . . . .	81
3.5.5	Une représentation des biomarqueurs fondée sur l'ontologie OGMS . . . . .	84
3.5.6	Le projet THESEUS MEDICO . . . . .	84
3.6	Discussion . . . . .	87
3.6.1	Problématique générale des contributions de la thèse . . . . .	87
3.6.2	Contributions de la thèse . . . . .	88
<b>4</b>	<b>Vers une Représentation Sémantique des Biomarqueurs d'Imagerie</b>	<b>91</b>
4.1	Introduction . . . . .	91
4.2	Contexte et objectif du travail . . . . .	92
4.3	Matériel et méthode . . . . .	93
4.3.1	Portée de l'ontologie des biomarqueurs d'imagerie . . . . .	93
4.3.2	Intégration des ontologies . . . . .	94
4.3.3	Réutilisation des ontologies en fonction des contextes . . . . .	99
4.3.4	Construction des modules ontologiques . . . . .	111
4.4	Résultat . . . . .	112
4.4.1	Développement de l'ontologie IBO . . . . .	112
4.4.2	Application de l'ontologie IBO pour la représentation de quelques biomarqueurs d'imagerie . . . . .	117
4.5	Discussion . . . . .	120
<b>5</b>	<b>Sémantique des données d'observation en neuro-imagerie selon un point de vue réaliste</b>	<b>123</b>
5.1	Introduction . . . . .	123
5.2	Contexte et objectif du travail . . . . .	124
5.2.1	Approche cognitive versus approche réaliste . . . . .	124
5.2.2	Sémantique des données d'observation en neuro-radiologie . . . . .	126
5.2.3	Objectif du travail . . . . .	127
5.3	Matériel et méthode . . . . .	127
5.3.1	Conception de l'ontologie VASARI . . . . .	128
5.3.2	Conception du travail expérimental . . . . .	128
5.4	Résultats . . . . .	130

---

5.4.1	Représentation réaliste et ontologique des résultats d'observation VASARI . . . . .	130
5.4.2	Expérimentations : annotation et exploitation sémantique des données d'observation VASARI . . . . .	134
5.5	Discussion . . . . .	142
<b>6</b>	<b>Discussion</b>	<b>149</b>
6.1	Introduction . . . . .	149
6.2	Rappel des objectifs et contributions de la thèse . . . . .	150
6.3	Évaluation de nos contributions . . . . .	150
6.3.1	Évaluation de l'ontologie IBO . . . . .	150
6.3.2	Évaluation de l'ontologie VASARI et des expérimentations . . . . .	151
6.4	Évaluation du travail par rapport à l'existant . . . . .	153
<b>7</b>	<b>Conclusion générale et perspectives</b>	<b>155</b>
	<b>Bibliographie</b>	<b>157</b>



# Table des figures

2.1	Comparaison entre les deux approches de prise en charge des patients : ancienne vs nouvelle. Les couleurs des patients indiquent leur diversité au niveau génétique. Source : PM presentation at the Healthcare Drinks networking event . . . . .	19
2.2	Définition du concept d'imagerie quantitative (Quantitative Imaging ou QI) d'après le groupe de la RSNA (Radiological Society of North America) TQI (Toward Quantitative Imaging) extraite de [Buckler 2011b] . . . . .	21
2.3	Exemples de caractéristiques numériques mesurées en imagerie quantitative . . . . .	22
2.4	Calcul du volume de la rate ( $387,3 \text{ cm}^3$ ) à partir d'image CT, source Projet MEDICO . . . . .	23
2.5	L'IRM d'un patient atteint de la Sclérose en Plaques. La flèche rouge indique une lésion détectée sur la séquence, source : Wikipédia . . . . .	25
2.6	Distribution des valeurs ADC au sein d'une tumeur d'une patiente atteinte d'un cancer du sein. L'échelle de couleur va du rouge (ADC élevé) au bleu (ADC faible) [Abramson 2015]. . . . .	26
2.7	Le spectre (b) montre un pic de la valeur du métabolite Myo-inositol (mI) de l'astrocytome localisé dans le lobe frontal droit (lésion cérébrale encadrée sur l'image pondérée T1 avec prise de contraste (a)) [Kao 2013]. . . . .	28
2.8	Utilisation du traceur $^{18}\text{F}$ -FDG pour surveiller la réponse thérapeutique en cancérologie. Cas clinique d'un patient avec Hodgkin Lymphome, évalué après huit séances de chimiothérapie : image CT (cadre à gauche), image TEP au $^{18}\text{F}$ -FDG (cadre à droite) et des images de fusion TEP/CT (cadre au centre). La région de concentration du traceur FDG correspond à la localisation d'un petit ganglion supraclaviculaire droit sur l'image CT, indiquant clairement une récurrente lymphome Hodgkin, extraite de [ESR 2010]. . . . .	30
2.9	Exemple d'évaluation de la progression d'une lésion ganglionnaire sur une TDM à quelques mois. Augmentation des valeurs des plus grands diamètres des lésions de 37,2mm à 51,7mm [Chan 2007]. . . . .	34

2.10	(a) L'image pondérée T1 contrastée représente un astrocytome de classe II selon la classification de l'Organisation Mondiale de la Santé (OMS) (bas grade). (b) La carte $K_{trans}$ montre aucune fuite du produit de contraste dans la région tumorale, ce qui suggère une faible valeur de $K_{trans}$ [Kao 2013]. . . . .	35
2.11	Processus de recherche et de développement des médicaments [FDA 2005]	38
2.12	Quelques exemples d'usage des biomarqueurs dans les différentes phases de R&D des médicaments, extraite de [FDA 2005] . . . . .	38
2.13	Structure standard d'un profil QIBA [Kevin 2016]. . . . .	40
2.14	Les étapes de mesure de la valeur SUV [RSNA 2007] . . . . .	42
2.15	Figure extraite de [ESR 2013] décrivant les principales étapes de développement d'un biomarqueur d'imagerie. . . . .	44
2.16	Les étapes d'évaluation des biomarqueurs d'imagerie d'après le "Institute of Medicine" [Downing 2001]. . . . .	45
2.17	Schéma conceptuel autour du concept grandeur. Les concepts encadrés sont les quelques notions clés qui sont abordées dans ce rapport. Pour mieux comprendre les différentes significations des relations de spécialisation entre les concepts il faut retourner à la source [Joint Committee for Guides in Metrology 2012]. . . . .	48
2.18	Rapport radiologique pour regrouper les données d'observation issues du diagnostic des patients de la SEP, extraite de [Filippi 2016] . . . . .	53
2.19	Exemple de rapport généré par le "Quantib Brain", extraite de <a href="http://www.quantib.com/">http://www.quantib.com/</a> . Le rapport contient des mesures de (1) volume de la matière grise, matière blanche, le liquide cérébro-spinal et du cerveau, (2) de l'hyper-intensité dans la matière blanche. L'objet de ce rapport est l'étude de l'évolution de ces mesures entre les deux années 2010 et 2014 pour le patient en question. . . . .	54
2.20	Exemple de rapport de MSmetrix, Dans la section "visual results" les points rouge représente les zones d'hypo-intensité et la couleur bleue représente la matière grise. Dans la section "brain volumes" "annual volume change" est le pourcentage de changement du volume du cerveau/ matière grise entre l'année 2015 et 2014. la valeur "normal volume change" représente le changement de volume annuel moyen pour une personne en bonne santé. . . . .	57
3.1	Structure globale de la terminologie SNOMED CT . . . . .	62
3.2	Structure globale de la terminologie Radlex . . . . .	63
3.3	CAD d'une mammographie, les résultats d'imagerie sont codés sous le format DICOM SR [Clunie 2007a] . . . . .	65
3.4	Les axes sémantiques de l'annotation des contenus informationnels des DICOM SRs, extraite de ce papier [Després 2007] . . . . .	67
3.5	Les groupes d'annotation du modèle AIM [Channin 2010] . . . . .	68
3.6	Le diagramme UML du modèle AIM [Channin 2010] . . . . .	69

---

3.7	L'infrastructure des technologies du Web sémantique, d'après T. Berners Lee . . . . .	71
3.8	Visualisation de l'ontologie Radlex sous le Bioportal NCBO . . . . .	74
3.9	Les différents niveaux de granularité des ontologies OBO couvrant des données de type "BFO:continuant" et "BFO:occurrent" . . . . .	76
3.10	Annotation avec le modèle OA d'une lésion de type glioblastome identifiée sur une séquence de type IRM [Ciccarese 2014] . . . . .	78
3.11	Extrait des types de Datasets les plus généraux basés sur (a) la modalité, (b) le traitement, (c) l'entité explorée et (d) les types de Dataset de niveau inférieur qui héritent des propriétés des trois axes sémantiques précédents. [Temal 2008] . . . . .	79
3.12	Extrait de la taxonomie des traitements qui portent sur les "Datasets", [Lynda 2008] . . . . .	80
3.13	Structure générale de l'ontologie QIBO, extraite de [Buckler 2013a] . . . . .	82
3.14	Les principaux concepts de l'ontologie BiomRKRS, extraite de [Ofoghi 2014] . . . . .	83
3.15	Définitions semi-formelles des principaux concepts du biomarqueur, extraite de [Ceusters 2015] . . . . .	85
3.16	Les principaux concepts OGMS utilisés, extraite de [Ceusters 2015] . . . . .	85
3.17	Structure de l'ontologie MEDICO [Seifert 2010] . . . . .	86
3.18	Représentation de la taille des ganglions lymphatiques dans le plan axial selon les critères RECIST avec le modèle MCI [Oberkampff 2012] . . . . .	87
4.1	Le template "TID 1410" nommé "Planar ROI measurements" structure les informations relatives aux mesures qui sont calculées à partir des régions d'intérêt planaires. Ces ROIs peuvent être référencées sur l'image à travers des formes de type "SCoord" ou bien des images de segmentation. . . . .	95
4.2	Les paramètres du concept "ROI measurements" du template "TID 1410" [Clunie 2012] . . . . .	96
4.3	Les types de l'entité "measurement objects" (tableau "CID RPH7601 measurement object types" [Clunie 2012]) . . . . .	96
4.4	La hiérarchie des classes de l'ontologie BFO [Smith 2012] . . . . .	97
4.5	Le problème de modélisation des processus de mesurage de qualités temporelles ("John's heart beating").(extraite de la présentation "BFO tutorial" de Barry Smith à la conférence icbo 2011) . . . . .	99
4.6	La structure générale de l'ontologie IBO . . . . .	100
4.7	La modélisation d'une investigation selon l'ontologie OBI [Smith 2015] . . . . .	101
4.8	Représentation du processus de mesurage de la glycémie avec l'ontologie OBI, [Bandrowski 2016] . . . . .	102
4.9	L'alignement de trois classes principales de l'ontologie IAO à l'ontologie BFO . . . . .	103
4.10	Classification générale des entités anatomiques dans l'ontologie FMA, extraite de la présentation de Barry Smith sur BFO2 . . . . .	105

4.11	Concepts extraits de l'ontologie PATO . . . . .	106
4.12	Schématisation de l'évolution temporelle de la maladie au sein d'un organisme de l'étape pré-clinique à l'étape de traitement. Toutes les définitions des termes sont données dans ce papier [Scheuermann 2009]	108
4.13	Les entités extraites de l'ontologie OGMS . . . . .	109
4.14	L'hierarchie générale de l'ontologie DHO . . . . .	112
4.15	Les principales composantes de type "BFO:continuant" de l'ontologie IBO. Les classes sont représentées avec des cadres et les relations entre les différentes classes sont dénotées par des flèches. . . . .	113
4.16	Les principales composantes de type "BFO:occurrent" de l'ontologie IBO. Les classes sont représentées avec des cadres et les relations entre les différentes classes sont dénotées par des flèches. . . . .	114
4.17	Représentation du processus de mesure de la valeur rCBV avec l'ontologie IBO : le cadre "a" décrit les entités qui sont impliquées dans le processus de préparation du sujet, le cadre "b" correspond au processus de création d'image et le cadre "c" détaille l'ensemble de données du processus de traitement. Nous notons que les noms des processus sont mis en gras, les classes qui ne sont pas des processus sont contenues dans des cadres et les relations sont désignées par des arcs.	118
4.18	La représentation du critère VASARI F <sub>29</sub> avec l'ontologie IBO. Nous notons que les processus sont représentés en gras, les classes qui ne sont pas des processus sont contenues dans des boîtes et les relations sont désignées par des arcs. . . . .	120
5.1	La capture d'écran montre un exemple de l'interface Clear Canvas utilisée pour la saisie des données de la terminologie VASARI [Gutman 2013] . . . . .	127
5.2	La structure générale de l'ontologie VASARI . . . . .	129
5.3	Le modèle de base des principales classes dans l'ontologie VASARI .	130
5.4	Hierarchie des classes de type "VASARI:cerebral tumor component" sous Protégé . . . . .	131
5.5	Exemple illustratif d'exploitation sémantique des connaissances anatomiques de l'ontologie FMA (sous l'outil protégé) ; les noms des instances sont générés automatiquement avec notre outil d'annotation, l'ID du patient 900_00_1961 est inclus pour désigner les entités existantes dans le corps du patient. . . . .	140
5.6	Résultat d'exécution d'une requête SPARQL sur les observations du neuro-radiologue 1 de la base REMBRANDT (démonstration avec l'outil CORESE). FMA_61823 désigne "FMA:lobe of cerebral hemisphere", FMA_61817 désigne le "FMA:cerebral hemisphere" et BFO_0000171 indique la relation "OBI:located_in_at_some_time". . . . .	147



---

5.7 Exemple d'une sortie de validation : détection d'une contradiction logique, IAO_000004 désigne la relation "IAO:has_measurement_value" et _paldmd_900_00_1961 fait référence à la valeur de mesure de l'axe perpendiculaire le plus long du patient 900_00_1961. . . . .	148
--	-----



# Liste des tableaux

2.1	Bases physiques et aspects mesurés des principales modalités d'imagerie	23
2.2	Réponses tumorales dans RECIST 1.1 . . . . .	34
2.3	Principales données d'imagerie en cancer d'après les statistiques du QIN [Levy 2012] . . . . .	50
3.1	Listing de différentes ressources du Biportal NCBO ordonnées par le nombre des classes . . . . .	77
4.1	Les trois niveaux de relations de l'ontologie RO de la fonderie OBO .	98
4.2	Alignement des termes de QIBO aux classes d'OBO . . . . .	109
4.3	Correspondance entre les termes génériques de QIBO et les termes d'IBO. . . . .	110
4.4	Les classes utilisées dans les cas d'utilisation relatifs au domaine des gliomes . . . . .	119
4.5	Description de quelques critères VASARI [TCGA-Research-Group 2014], la valeur "n/a" ("not applicable") indique que le critère n'est pas applicable au cas étudié et la valeur "none" est attribuée quand l'objet d'étude du critère n'existe pas ou bien n'est pas observable sur la séquence d'image. . . . .	121
5.1	Résumé de la nature des problèmes de modélisation rencontrés dans la description sémantique des critères VASARI . . . . .	145
5.2	Description des critères VASARI sous forme d'entités et de relations ontologiques . . . . .	146



# Chapitre 1

## Introduction Générale

L'image médicale contient des informations pertinentes pour aider les médecins dans l'étude de la pathologie dont souffre le patient et choisir l'approche thérapeutique la plus adaptée (médicament, chirurgie, radiothérapie, etc.). Des techniques d'acquisition et d'analyse d'images ont émergé pour l'extraction des informations quantitatives afin d'affiner l'exploration radiologique des organes et des tissus explorés et ne pas se contenter de la simple interprétation visuelle des images. Ces nouvelles techniques d'imagerie quantitative sont entrain de modifier en profondeur le diagnostic médical et l'emploi des mesures radiologiques. Dans la littérature, ces mesures sont connues sous le nom de biomarqueur d'imagerie.

Pour simplifier, le concept de biomarqueur d'imagerie fait référence aux mesures radiologiques qui caractérisent les processus biologiques chez les sujets et qui aident les cliniciens en particulier dans le diagnostic précoce des pathologies et l'évaluation des réponses thérapeutiques, par exemple, la mesure de la taille d'une lésion, le volume d'une tumeur, la perfusion sanguine dans une région d'intérêt, etc. Les biomarqueurs d'imagerie sont devenus incontestablement l'un des éléments clefs non seulement dans le cadre de la prise en charge clinique des patients mais aussi dans les essais cliniques visant à évaluer des nouvelles thérapies, de façon à améliorer leur efficacité et diminuer leur coût.

L'augmentation du nombre de travaux de recherche réalisés autour des biomarqueurs d'imagerie et l'utilisation croissante de ces mesures dans la routine radiologique souligne la nécessité de la mise en place d'outils de gestion performants. L'utilisation de technologies sémantiques nous semble pertinente pour bien supporter les activités comme la recherche efficace des mesures les plus adaptées à un problème de quantification précis, l'interopérabilité dans le domaine biomédical, la comparaison entre différents jeux de données, et la mise en œuvre des études multi-centriques, etc.

Ainsi, cette thèse a pour but de proposer une représentation sémantique des entités les plus importantes du domaine des biomarqueurs d'imagerie couvrant la représentation des données de provenance des mesures. Avec le standard DICOM (format d'échange d'images et de comptes rendus archivage très répandu dans la pratique clinique) et le système d'archivage et de communication d'images PACS

(Picture Archiving and Communication System), on peut inclure que des données d'imagerie qui décrivent l'anatomie, les modalités d'image, les régions d'intérêt étudiées mais on ne peut pas inclure des informations sur les traitements d'imagerie exécutés, les maladies étudiées, les méthodes de calcul des biomarqueurs d'imagerie, etc. Ainsi, le lien entre le biomarqueur et les images à partir desquelles le biomarqueur a été calculé peut être exprimé mais les méta-données supplémentaires sur les méthodes d'imagerie, d'extraction des critères et d'analyse appliquées ne peuvent pas être spécifiées.

De notre point de vue, l'emploi des annotations sémantiques va permettre d'explicitier d'une manière formelle la sémantique des mesures et des observations radiologiques, en utilisant les termes, les relations ainsi que les axiomes d'une ontologie. Ceci permet de décrire et partager les informations relatives à cette ressource telle que le protocole de mesurage, les traitements d'imagerie, le contexte clinique, ou encore les processus de prise en charge, etc.

Cette thèse a deux contributions :

La première vise à formaliser la description des biomarqueurs d'imagerie à travers la définition d'une ontologie générique. L'ontologie résultante, respecte les bonnes pratiques de l'ingénierie ontologique : la réutilisation d'ontologies existantes, l'articulation autour d'une ontologie fondatrice, la modularité de ses composantes, la réutilisabilité dans d'autres contextes, etc. Ce travail a donné lieu à une présentation à la Conférence Semantic Web And Tools for Life Science (SWAT4LS) à Cambridge (UK) en 2015 [Amdouni 2015], puis aux Journées Francophones sur les Ontologies (JFO) à Bordeaux en 2016 [Amdouni 2016b], où ce travail a reçu le Prix de la meilleure présentation étudiante. Une version étendue de l'article présenté aux JFO est soumise et elle est en révision au "Journal of Data Semantics", qui doit publier une sélection des articles présentés aux JFO en 2016. Enfin, une troisième présentation<sup>1</sup> a été faite au Congrès Européen de Radiologie à Vienne (Autriche) en 2016, orientée sur l'application du modèle générique à plusieurs biomarqueurs utilisés en cancérologie (gliomes cérébraux).

La deuxième contribution propose une modélisation réaliste et ontologique de la sémantique des résultats d'observations des glioblastomes et elle est complétée par un travail d'implémentation de preuve de concept autour de la base REMBRANDT. Ce second travail a donné lieu à une présentation à la Conférence Knowledge Engineering and Ontology Development (KEOD) à Porto (Portugal) en 2016 [Amdouni 2016a]. Une version étendue en journal de cet article est en préparation.

Ce rapport de thèse est organisé comme suit :

- Le chapitre 2 introduit quelques notions de base liées aux techniques d'imagerie médicale et au concept de biomarqueur d'imagerie, et il décrit le contexte de la thèse.
- Le chapitre 3 présente les formats informatiques existants pour le partage de la sémantique des données quantitatives d'imagerie. A la fin de ce chapitre, nous décrivons brièvement les deux contributions de la thèse.

---

1. [www.myESR.org/epos](http://www.myESR.org/epos)

- Le chapitre 4 aborde le sujet de la représentation sémantique du concept de biomarqueur d'imagerie et introduit notre modèle ontologique en démontrant en quoi il aide à lever les ambiguïtés qui existent autour de ce concept.
- Le chapitre 5 traite le sujet de la modélisation ontologique et réaliste des données d'observation et montre la valeur ajoutée de l'adoption d'une telle approche dans le domaine de la neuro-imagerie.
- Le chapitre 6 propose une discussion générale et évalue la valeur ajoutée des contributions que nous proposons dans le cadre de cette thèse vis-à-vis d'autres projets existants.
- Enfin le chapitre 7 rappelle l'importance de l'enrichissement sémantique des données relatives aux biomarqueurs d'imagerie dans la mise en place des bio-banques d'images et présente quelques perspectives de notre travail.





## Chapitre 2

# Contexte général du travail de la thèse

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>18</b>
<b>2.2</b>	<b>L'avènement de la médecine de précision</b>	<b>18</b>
2.2.1	Les biomarqueurs, la stratification et les traitements appropriés au cœur de la médecine personnalisée	19
2.2.2	Promesses versus réalités de la médecine personnalisée	20
<b>2.3</b>	<b>L'imagerie quantitative au service de la médecine de précision</b>	<b>20</b>
2.3.1	Tomodensitométrie	21
2.3.2	Imagerie par Résonance Magnétique	23
2.3.3	Médecine nucléaire	29
<b>2.4</b>	<b>L'ère des biomarqueurs d'imagerie</b>	<b>31</b>
2.4.1	Définition du concept de biomarqueur d'imagerie	31
2.4.2	Utilité des biomarqueurs d'imagerie	32
2.4.3	Standardisation des biomarqueurs d'imagerie	39
2.4.4	Processus de développement des biomarqueurs d'imagerie	41
2.4.5	Classification des biomarqueurs d'imagerie	45
<b>2.5</b>	<b>Besoin d'intégration des données d'imagerie pour supporter le développement des bio-banques d'images</b>	<b>47</b>
<b>2.6</b>	<b>Contexte d'étude : représentation des observations d'imagerie médicale</b>	<b>50</b>
2.6.1	Description des besoins médicaux dans les applications dédiées à la délivrance de soins	50
2.6.2	Description des besoins médicaux dans les applications dédiées à la recherche	54

---

## 2.1 Introduction

L'objet d'étude de cette thèse concerne le partage des données quantitatives et qualitatives en imagerie médicale, afin de favoriser leur réutilisation dans la communauté exploitant les données d'imagerie. Dans ce chapitre, nous définissons plus en détail le contexte général de la thèse. Ce chapitre est organisé comme suit : la Section 2.2 présente le concept de la médecine de précision et ses différents avantages en pratique clinique. Dans la Section 2.3, nous passons en revue les différentes techniques d'imagerie médicale employées pour l'exploration des caractéristiques biologiques chez l'Homme. La Section 2.4 est dédiée à la présentation du domaine des biomarqueurs d'imagerie (définition du concept, utilité clinique, développement, etc.). Dans la Section 2.5, nous abordons la question de l'intégration des données d'imagerie dans le domaine biomédical. La section 2.6 présente quelques besoins fonctionnels autour de la problématique du partage des données d'imagerie à la fois en recherche et en pratique clinique.

## 2.2 L'avènement de la médecine de précision

Les individus sont différents les uns des autres sur le plan biologique. Il existe plusieurs formes et sous-types d'une même maladie possédant chacune leurs propres signatures moléculaires. Auparavant, pour une même maladie on adoptait une approche uniformisée pour tous les patients qui peuvent bénéficier d'un traitement. Les chimiothérapies, par exemple étaient d'une efficacité très variable dépendant des spécificités génétiques de chaque patient. Aujourd'hui, la compréhension précise des différents sous-types d'une même maladie permet d'adapter la prise en charge et de mettre en œuvre une thérapie ciblée. Dans une thérapie ciblée en cancérologie par exemple, les molécules d'un traitement spécifique s'attaquent en priorité aux cellules cancéreuses en ciblant les altérations génétiques présentes dans les tumeurs en fonction du profil biologique du patient et en fonction des caractéristiques moléculaires de sa maladie. Cette nouvelle pratique médicale est possible grâce à la *médecine personnalisée*.

La médecine personnalisée connue aussi sous le nom de la *médecine de précision*, est basée sur l'adaptation des traitements médicaux aux caractéristiques individuelles de chaque patient durant les processus de délivrance des soins. La médecine personnalisée est apparue en 2010 et elle a comme principal objectif de fournir « le bon traitement avec la bonne dose au bon patient et au bon moment » [ESR 2011] [Forward 2012]. En médecine de précision, le profil génétique d'un patient est la clef de voute dans le processus décisionnel ; il permet d'orienter le choix vers des thérapies ou des protocoles de traitement qui induisent la meilleure efficacité et le moins d'effets secondaires (voir Figure 2.1).

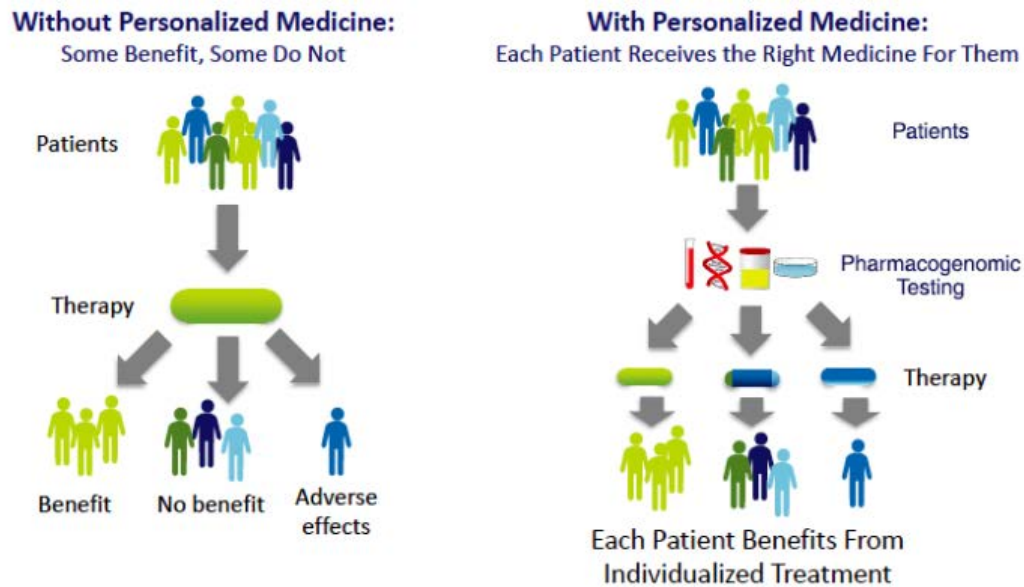


FIGURE 2.1 – Comparaison entre les deux approches de prise en charge des patients : ancienne vs nouvelle. Les couleurs des patients indiquent leur diversité au niveau génétique. Source : PM presentation at the Healthcare Drinks networking event

### 2.2.1 Les biomarqueurs, la stratification et les traitements appropriés au cœur de la médecine personnalisée

La précision de la médecine personnalisée provient notamment de son emploi des *biomarqueurs* (c'est à dire marqueurs de processus biologiques du patient) pour aider à sélectionner un traitement adapté au patient. Par exemple, en cancérologie, des tests diagnostiques doivent être effectués pour permettre la détermination des *profils moléculaires* des tumeurs (profils décrivant les modifications caractérisant la tumeur) et proposer au patient une thérapie ciblant les gènes altérés. On note que les biomarqueurs renseignent d'une façon fiable sur la gravité de la tumeur (marqueur pronostique) ou bien l'efficacité ou la tolérance à un traitement donné (marqueur prédictif). Les biomarqueurs prédictifs permettent de prédire si le médicament peut être sans danger et efficace, sans danger et inefficace, dangereux et inefficace ou dangereux et efficace pour le patient.

Ces deux notions de biomarqueur et de profil moléculaire, permettent la stratification des patients c'est à dire la catégorisation des malades en fonction de leur profils moléculaires pour traiter les sous-groupes (idéalement chaque patient) d'une façon plus efficace [Crommelin 2011]. Ainsi, la médecine personnalisée est fondamentalement basée sur cette association entre le profil biologique du patient et les caractéristiques moléculaires de sa maladie.

### 2.2.2 Promesses versus réalités de la médecine personnalisée

En réalité, le nombre de biomarqueurs employés dans la pratique clinique pour des maladies comme le cancer, la maladie d'Alzheimer, l'asthme, l'autisme, l'hépatite, etc est réduit. Le frein majeur à cette croissance est du à l'hétérogénéité des pathologies qui rend difficile la personnalisation des soins délivrés. Cette diversité de pathologies concerne l'aspect phénotypique, moléculaire et clinique :

- **Diversité phénotypique** : elle concerne la variabilité des structures pathologiques au plan morphologique (taille, forme, contour, texture, etc.), fonctionnel (diffusion, perfusion, élasticité, etc.) et moléculaire (activité métabolique, expression des gènes, etc.) sur les séquences d'images.
- **Diversité moléculaire** : il s'agit de l'hétérogénéité génomique des tumeurs qui est due au fait que différents gènes peuvent intervenir dans le développement de la même pathologie.
- **Diversité clinique** : elle est due à la variabilité de la réponse thérapeutique chez les patients.

Pour améliorer l'application du concept de la médecine personnalisée à différentes spécialités médicales, des efforts doivent être menés pour supporter le développement et l'intégration des bases de données biologiques (séquençage du génome humain), des méthodes de caractérisation des patients et des outils d'analyse de grandes masses de données, etc. En imagerie médicale, des techniques d'imagerie avancées sont entrain d'émerger pour réaliser des analyses à l'échelle moléculaire afin d'améliorer la qualité des applications thérapeutiques proposées aux patients. Dans cette thèse nous sommes intéressés aux paramètres d'imagerie caractérisant les patients en médecine de précision, ces paramètres sont connus sous le nom des *biomarqueurs d'imagerie*.

## 2.3 L'imagerie quantitative au service de la médecine de précision

L'imagerie quantitative fait référence à l'extraction et l'utilisation de caractéristiques numériques extraites des images médicales dans le cadre de la recherche et de la pratique clinique [Kessler 2015] (voir Figure 2.3). L'objectif de l'imagerie quantitative est de faire évoluer les techniques d'imagerie médicale pour améliorer la compréhension des pathologies à l'échelle moléculaire à travers l'emploi des nouvelles méthodes numériques d'acquisition d'image, de traitement, d'archivage et d'affichage des résultats [Buckler 2013b] [Jaffer 2005]. Les données issues de l'imagerie quantitative principalement les techniques de l'*Imagerie par Résonance Magnétique* (IRM) et celles de la *médecine nucléaire* sont capables de fournir des informations plus détaillées et plus précises sur le fonctionnement de l'objet étudié [ESR 2010] [Buckler 2011b]. Cette évolution spectaculaire des techniques d'imagerie doit permettre l'amélioration de la compréhension des phénomènes pathologiques, et au delà, l'adoption des biomarqueurs d'imagerie doit permettre d'assister les médecins dans la prise de décision thérapeutique appropriée.

Quantitative imaging is the extraction of quantifiable features from medical images for the assessment of normal or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal. Quantitative imaging includes the development, standardization, and optimization of anatomical, functional, and molecular imaging acquisition protocols, data analyses, display methods, and reporting structures. These features permit the validation of accurately and precisely obtained image-derived metrics with anatomically and physiologically relevant parameters, including treatment response and outcome, and the use of such metrics in research and patient care.

FIGURE 2.2 – Définition du concept d'imagerie quantitative (Quantitative Imaging ou QI) d'après le groupe de la RSNA (Radiological Society of North America) TQI (Toward Quantitative Imaging) extraite de [Buckler 2011b]

Les techniques d'imagerie médicale sont réparties en méthodes d'imagerie avec radiations ionisantes et d'autres sans radiations ionisantes. Les méthodes d'imagerie avec radiations ionisantes sont basées sur les rayons X (première image par rayons X apparue en 1896) et les isotopes radioactifs (médecine nucléaire apparue en 1952). Les rayons X traversent le corps et c'est leur atténuation par les os, les organes, les tissus (musculaires, nerveux, etc.) et les vaisseaux qui permet d'obtenir une image radiologique ou *tomodensitométrie*. En médecine nucléaire, les rayonnements sont émis par le corps après injection d'un produit faiblement radioactif nommé traceur ou radiopharmaceutique. Les principaux examens sont la scintigraphie et la *Tomographie par Émission de Positons*. Un appareil capte l'émission des rayons gamma et reconstitue une image des zones où s'est fixé le traceur.

L'imagerie par Résonance Magnétique et la SRM *Spectrométrie par Résonance Magnétique* (SRM) (découverte en 1946) sont des méthodes d'imagerie sans radiations ionisantes mais des champs de radio-fréquences. Les appareils d'IRM n'émettent aucun rayonnement ionisant ; cette technique repose sur le phénomène de résonance des atomes d'hydrogène du corps quand ils sont soumis à un champ magnétique. La machine fonctionne avec un puissant électro-aimant et produit des images en coupe du corps, en deux ou trois dimensions. Le tableau 2.1 résume les bases physiques des différentes modalités d'imagerie et les phénomènes biologiques mesurés.

### 2.3.1 Tomodensitométrie

La tomodensitométrie plus couramment appelée scanner (en anglais Computerized Tomography (CT)) est une technique d'imagerie médicale qui est basée sur la mesure de la quantité de rayons X absorbés par les tissus étudiés qui sont de densités

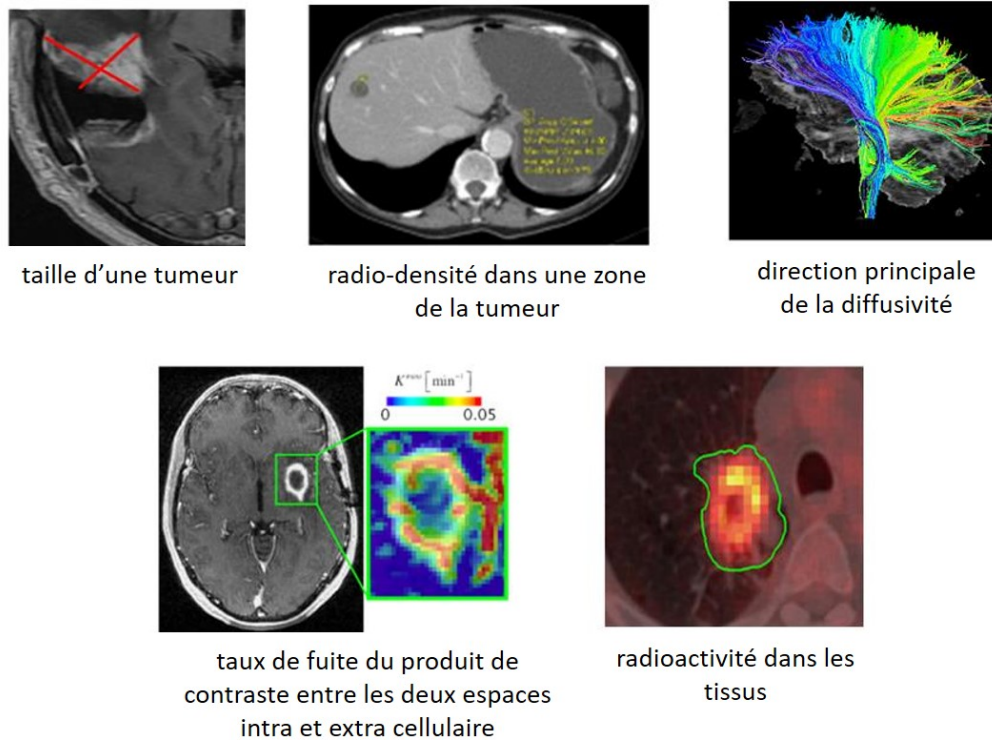


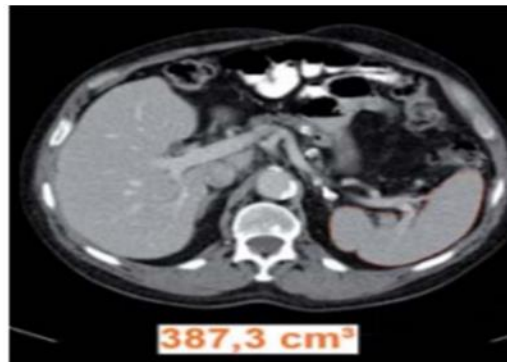
FIGURE 2.3 – Exemples de caractéristiques numériques mesurées en imagerie quantitative

différentes. Pour l'acquisition de ce type d'image, un appareil d'imagerie émet les rayons X et mesure les rayonnements ayant traversé les tissus sous des angles multiples afin de produire des séquences d'images. Des coupes bidimensionnelles (2D) sont d'abord générées puis elles sont rassemblées pour produire un jeu de données en trois dimensions (3D). On note que, on peut avoir recours à l'injection d'un produit de contraste dans les examens CT pour permettre de mettre encore plus en évidence certaines parties qui sont difficiles à voir comme par exemple les vaisseaux sanguins. D'une façon générale en imagerie médicale, pour mieux visualiser un organe, une tumeur ou une partie du corps, il est parfois nécessaire d'utiliser un produit de contraste. Un produit de contraste a la propriété de fortement absorber les rayons X ou de radiofréquences et donc de rendre très visibles les tissus où ce produit est présent (qui apparaissent alors hyperdenses, c'est-à-dire plus « blancs » sur l'image), il peut être administré au patient par voie orale, par injection intraveineuse ou intra-articulaire, etc.

En pratique clinique, les examens CT sont très utiles pour la localisation avec précision des endroits où l'intensité de signal change et ainsi faciliter la détection des lésions internes au sein de certains organes, des métastases, l'observation de la circulation sanguine, etc. La quantification du volume tumoral à partir des séquences

Modalité d'imagerie	Base physique	Aspect biologique
Tomodensitométrie	rayons X	densité des tissus
Imagerie par Résonance Magnétique	radiofréquence	distribution anatomique de l'eau et de la graisse dans les organes
Médecine nucléaire	rayons gamma	nombre de désintégrations dans chaque pixel/voxel d'image

TABLE 2.1 – Bases physiques et aspects mesurés des principales modalités d'imagerie

FIGURE 2.4 – Calcul du volume de la rate (387,3 cm<sup>3</sup>) à partir d'image CT, source Projet MEDICO

CT (Figure 2.4) permet d'estimer la croissance tumorale, cette mesure peut être utilisée dans les essais cliniques en cancer de poumons par exemple pour classer les patients en fonction de leurs réponses (sensibilités ou résistances) à une thérapie ciblée [Zhao 2010] ou bien pour le suivi de leurs réponses thérapeutiques dans les essais cliniques [Eisenhauer 2009].

### 2.3.2 Imagerie par Résonance Magnétique

A partir des années 1970, la *Résonance Magnétique Nucléaire* (RMN) [Purcell 1946] a été appliquée à la médecine. La RMN est utilisée pour déterminer la composition chimique d'un échantillon, le volume à explorer est placé dans un champ magnétique puissant et des ondes de radio-fréquence lui sont appliquées. L'onde de radio-fréquence est une onde électro-magnétique dont la fréquence est choisie pour permettre le phénomène de résonance. Le principe de l'IRM qui est basé sur la RMN, repose sur l'étude de la nature des tissus constituant le corps humain en chaque voxel. Au cours d'un examen IRM, la machine envoie différents signaux de radio-fréquence afin de faire entrer les protons (atomes d'hydrogène) en résonance et modifie le vecteur d'aimantation (l'axe de rotation du proton). La magnétisa-

tion longitudinale du proton devient transversale. C'est le phénomène d'excitation. Après l'application des ondes de radio-fréquence, les protons reviennent à leur état initial (magnétisation initiale) et en fonction de leur propriété, de leur voisinage ils émettront à leur tour certains signaux électromagnétiques, ce seront ces derniers qui seront à l'origine des images générées par l'IRM.

En d'autres termes, l'IRM est en mesure de fournir des images en fonction de la densité des tissus en atomes d'hydrogène (densité protonique). D'une manière simplifiée, les images IRM sont des cartographies de la distribution anatomique de l'eau et de la graisse dans les organes. On souligne qu'en RMN, la pondération de l'image (c'est à dire relaxation longitudinale appelé temps T1 et relaxation transversale appelé temps T2) des différents tissus d'un organisme varie en fonction des paramètres d'acquisition IRM. Ces paramètres sont le temps de répétition entre deux excitations et le temps d'écho, temps entre le signal d'excitation et la réception de l'écho. Le signal liquidien est défini par un hypersignal T2 et un hyposignal T1 et le signal graisseux apparaît en hypersignal T1 et hypersignal T2 [Haacke 1999]. En pratique clinique, la technique d'imagerie IRM a été particulièrement pertinente dans la production d'images du cerveau, en raison de sa capacité à enregistrer des signaux qui peuvent distinguer les tissus mous (la matière grise et la matière blanche).

Il existe plusieurs types d'IRM, chaque type explore un aspect biologique particulier : l'IRM morphologique appelée aussi *IRM conventionnelle* qui permet d'étudier la morphologie de la partie du corps explorée, l'*IRM dynamique* qui étudie la micro-vascularisation au sein des tumeurs, l'*IRM fonctionnelle* qui teste les fonctions cérébrales, et la *Spectrométrie par Résonance Magnétique Nucléaire* qui est basée sur la spectroscopie par la RMN et permet d'étudier la biochimie des tissus et organes sans avoir recours à une biopsie ou chirurgie.

### 2.3.2.1 IRM conventionnelle

L'IRM conventionnelle ou encore l'IRM anatomique forme l'image médicale à partir de la distribution bi-dimensionnelle des molécules ( $H_2O$ ) qui se trouvent à l'intérieur d'un organe/tissu. Il existe plusieurs séquences d'IRM, principalement : les séquences en pondération T1, T2, FLAIR (Fluid Attenuated Inversion Recovery) et séquences d'IRM pondérées T1 après injection de Gadolinium.

Les techniques d'IRM permettent la réalisation de plusieurs mesures de biomarqueurs d'imagerie, entre autres le volume cérébral. Pour le diagnostic et le suivi des patients atteints de sclérose en plaques par exemple, les neuro-radiologues ont dans la plupart de temps recours à des séquences d'IRM de type T1, T1 après injection de Gadolinium et FLAIR. Ces séquences permettent l'évaluation de l'atrophie du volume du tissu cérébral [Bjartmar 2000] et la détection de l'apparition de nouvelles lésions chez les patients de la sclérose en plaques [Giorgio 2013] [Popescu 2013]. La Figure 2.5 montre une image IRM d'un patient atteint de sclérose en plaques.

**A. Séquence FLAIR :** La séquence FLAIR est généralement pondérée T2 (substance blanche plus noire que le cortex) et elle est très utilisée en imagerie cérébrale.



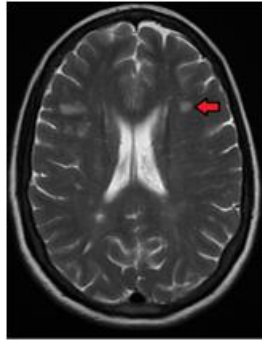


FIGURE 2.5 – L'IRM d'un patient atteint de la Sclérose en Plaques. La flèche rouge indique une lésion détectée sur la séquence, source : Wikipédia

Elle est caractérisée par la suppression du signal provenant du liquide céphalo-rachidien et donc la maximisation du contraste avec les lésions voisines. Cette séquence est plus sensible que la séquence T2 aux effets paramagnétiques du gadolinium après injection. La séquence FLAIR est très utilisée en SEP pour détecter les lésions en hypersignal autour des ventricules cérébraux, dans la moelle et dans la fosse postérieure (en SEP, les zones très blanches sur les séquence FLAIR sont visualisées en iso- ou en hyposignal en T1). Ces lésions sont de taille variable, de forme arrondie, sans effet de masse sur les autres structures.

**B. Séquence IRM pondérée T1 :** Le but de la séquence pondérée en T1 est de mettre en place les différences de temps de relaxation T1 entre les tissus. Sur cette séquence, les lésions apparaissent en signal hypointense dans la substance blanche. Toutefois cette séquence est moins sensible à la mise en évidence de ces anomalies macroscopiques. La séquence d'IRM pondérée T1 est utilisée dans plusieurs maladies comme par exemple, dans la maladie d'Alzheimer. Les patients atteints de la maladie d'Alzheimer sont notamment diagnostiqués grâce aux IRM pondérées en T1 qui permettent la détection de l'atrophie hippocampique qui reflète la perte neuronale [Zarow 2005].

### 2.3.2.2 IRM non conventionnelle

L'IRM non conventionnelle permet d'obtenir des informations morphologiques, biophysiques, métaboliques ou fonctionnelles et de mieux étudier les phénomènes physiopathologiques à travers des marqueurs d'imagerie ; des structures pathologiques d'apparence normale en IRM conventionnelle peuvent être mises en évidence en IRM non conventionnelle. Ces techniques d'IRM nécessitent des appareils plus avancés, des protocoles d'acquisition plus complexes, et des procédures de post-traitements et d'analyse plus sophistiquées. Parmi ces techniques on peut citer l'IRM de diffusion et l'imagerie du tenseur de diffusion, l'IRM de perfusion, la Spectroscopie par Résonance Magnétique et l'IRM fonctionnelle.

**A. IRM de diffusion et imagerie du tenseur de diffusion :** Les séquences d'IRM de diffusion appelée aussi DWI (Diffusion Weighted Imaging) explorent le mouvement aléatoire de l'eau, ce phénomène de diffusion résulte de la mobilité aléatoire de ces molécules (appelé mouvement Brownien) et il est dirigé suivant l'architecture des tissus biologiques [Le Bihan 1986] [Le Bihan 2003]. Les images de diffusion rendent disponibles des paramètres quantifiables telle le Coefficient Apparent de Diffusion (ADC). Ce coefficient est calculé automatiquement par les logiciels informatiques qui effectuent une comparaison entre les deux images : image pondérée en T2 et image pondérée en diffusion.

En cancérologie, la diffusion des molécules d'eau dépend indirectement de la cellularité du milieu. Plus un tissu est riche en cellules, plus la diffusion des molécules d'eau sera restreinte. Les tissus cancéreux étant constitués d'une prolifération anarchique de cellules, la diffusion de l'eau au sein de ces tissus est habituellement réduite. L'IRM de diffusion a montré son efficacité dans le cancer du sein [Yankeelov 2007] (Figure 2.6), le cancer de la prostate [Woodfield 2010] [Verma 2011] et le cancer du cerveau pour la détection de la mort cellulaire [Schmainda 2012], etc.

L'estimation de l'architecture locale des tissus peut aussi être employée pour le diagnostic dans différentes catégories de pathologies cérébrales : tumorales (lymphome cérébral, kyste épidermoïde et cholestéatome), infectieuses (abcès cérébral à pyogènes, encéphalite herpétique), dégénératives, inflammatoires (sclérose en plaques), traumatiques (choc, fracture) [Schaefer 2000] [Ron 2003].

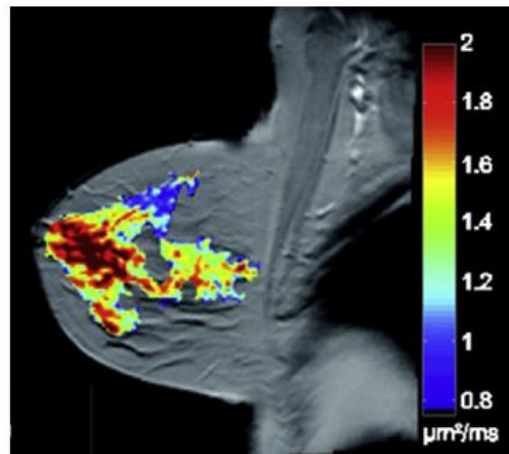


FIGURE 2.6 – Distribution des valeurs ADC au sein d'une tumeur d'une patiente atteinte d'un cancer du sein. L'échelle de couleur va du rouge (ADC élevé) au bleu (ADC faible) [Abramson 2015].

La technique de tenseur de diffusion reflète l'anisotropie [Le Bihan 2001]. Cette séquence permet d'obtenir des cartographies d'ADC [Le Bihan 2013] et d'Anisotropie Fractionnelle (FA) [Basser 2011]. En neuro-radiologie, des logiciels d'imagerie permettent de reconstruire les principaux faisceaux de fibres de la substance blanche (tractographie). Une des applications majeures consiste à visualiser les principaux

faisceaux de substance blanche au voisinage d'une tumeur, pour évaluer s'ils sont envahis ou simplement refoulés et pour étudier la connectivité anatomique du cerveau (sain et pathologique) [Craddock 2013]. La modélisation mathématique de la diffusion des molécules d'eau par un tenseur permet de déterminer la direction prépondérante de la diffusion moyennée à l'échelle de voxel. À partir de ces paramètres quantitatifs issus des acquisitions IRM pondérées en diffusion, l'application d'algorithmes mathématiques permet de calculer la cohérence directionnelle de la diffusion des molécules d'eau appartenant à des voxels voisins et ainsi de visualiser artificiellement le trajet des fibres de substance blanche.

**B. IRM de perfusion :** L'IRM de perfusion connue aussi sous le nom de DCE-MRI (Dynamic Contrast Enhanced MRI) donne accès à des informations sur la micro-circulation capillaire des tissus [Tofts 2010]. Les paramètres quantitatifs essentiels [Tofts 1999] qu'elle évalue sont les volumes sanguins et des données temporelles (temps de transit, temps jusqu'au pic de contraste, etc.). L'objectif final de l'IRM de perfusion est de mesurer le débit sanguin qui irrigue l'organe exploré qui est exprimé en ml/min/100g afin d'estimer la perfusion tissulaire. La différenciation des tissus perfusés et non perfusés repose sur l'utilisation d'un marqueur intra-vasculaire de type soit exogène ou endogène (marqueur des noyaux d'hydrogène, ce marqueur est diffusible c'est-à-dire qu'il met en évidence les échanges entre les deux espaces vasculaire et extra-vasculaire).

La technique d'IRM de perfusion est employée en cancérologie pour calculer des biomarqueurs d'imagerie pronostiques, par exemple dans le cancer du sein [Ah-See 2008] ou dans le cancer de la prostate [Vos 2013] ainsi que des biomarqueurs d'imagerie prédictifs de la réponse tumorale après un traitement thérapeutique [Zahra 2007].

**C. Spectroscopie par Résonance Magnétique :** La Spectroscopie par Résonance Magnétique (SRM) [Harris 1986] est basée sur l'imagerie du proton de l'atome d'hydrogène ( $^1\text{H}$ ) elle permet de réaliser des explorations métaboliques du cerveau via la quantification des signaux de la RMN. Les fréquences de résonance des noyaux (protons) des métabolites cérébraux sont différentes en fonction de leur environnement chimique. Le signal RMN des protons alors obtenu est un spectre dont chaque pic fournit des renseignements sur la fréquence de résonance et l'amplitude ainsi la concentration du métabolite en question. Les principaux métabolites détectables sont le N-Acétyl-Aspartate (NAA) marqueur de l'intégrité ou du fonctionnement axonal ; la Choline (Cho) marqueur de métabolisme phospholipidique et de densité cellulaire ; la Créatine (Cr) marqueur de métabolisme énergétique cellulaire, le Myo-inositol (Ins ou mI) marqueur de l'inflammation et de la gliose, le Glutamate (Glu) neurotransmetteur, le Lactate (Lac) indicateur de souffrance cellulaire liée à une ischémie, une hypoxie ou une réaction inflammatoire. L'avantage de la spectroscopie est qu'elle permet de détecter précocement des anomalies métaboliques entraînant une souffrance cérébrale avant même de visualiser une anomalie en IRM classique,

ou d'orienter vers la nature d'un processus pathologique en question. La Figure 2.7 montre l'utilité de la mesure du métabolite Myo-inositol pour la détermination du grade de l'astrocytome et le guidage du geste chirurgical chez les patients atteints d'un cancer du cerveau [Kao 2013].

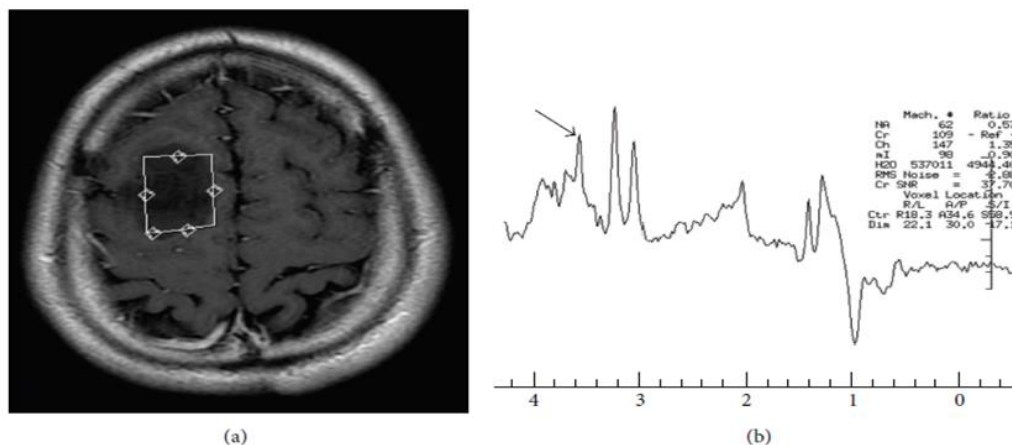


FIGURE 2.7 – Le spectre (b) montre un pic de la valeur du métabolite Myo-inositol (mI) de l'astrocytome localisé dans le lobe frontal droit (lésion cérébrale encadrée sur l'image pondérée T1 avec prise de contraste (a)) [Kao 2013].

**D. IRM fonctionnelle :** les séquences d'IRM fonctionnelle (IRMf) [Huettel 2004] permettent la détermination des régions cérébrales impliquées dans la réalisation d'une tâche (motrice, visuelle, cognitive, etc.). La réponse neuro-vasculaire se manifeste par une augmentation du débit sanguin cérébral dans les zones activées qui induit une augmentation de l'oxygénation du sang de la zone activée. Cette réponse hémodynamique locale correspond à l'effet BOLD (Blood Oxygenation Level Dependent Contrast) qui repose sur la différence intrinsèque de contraste du sang capillaire entre l'activation et le repos. Dans les zones activées, l'augmentation de la consommation d'oxygène est surcompensée par une large augmentation du flux sanguin. Pour l'interprétation des scans d'un seul patient, les neuro-radiologues ont recours à des méthodes statistiques qui permettent de faire de comparaison des données du patient avec une base de référence regroupant les résultats d'autres patients.

Le plus souvent cet examen est demandé dans le cadre du bilan d'évaluation pré-chirurgicale pour avoir plus d'information sur la localisation de la lésion par rapport aux aires du langage ou de la motricité. L'IRMf est aussi largement employée dans les études de la connectivité fonctionnelle principalement dans le domaine de la psychiatrie.

**2.3.3 Médecine nucléaire**

La médecine nucléaire est une technique d'imagerie fonctionnelle qui repose sur l'administration d'un traceur radioactif qui se distribue dans l'organisme en fonction de ses propriétés spécifiques. Ces traceurs radioactifs sont composés d'un produit physiologique et d'un isotope radioactif. On étudie la répartition du traceur en repérant sa désintégration grâce à l'émission de particules (photon ou positon). Des dispositifs détecteurs détectent les rayonnements émis par les isotopes radioactifs et déterminent la zone où a eu lieu la désintégration, c'est cette information quantitative que l'on représente sous la forme d'une image (scintigraphie ou/et Tomographie d'Émission de Positons (TEP)) faisant apparaître en couleurs les zones de forte concentration du traceur. Les principaux radio-traceurs utilisés en TEP sont le  $^{18}\text{F}$ -FDG ( $^{18}\text{F}$ -fluorodeoxyglucose), le  $^{18}\text{F}$ -Choline ( $^{18}\text{F}$ fluorocholine) et le F-DOPA (Fluorodopa).

L'emploi de la médecine nucléaire dans la pratique clinique a un grand intérêt dans l'exploration fonctionnelle et métabolique des tissus afin de permettre par exemple la détection fine de lésions profondes, le suivi thérapeutique de nombreuses pathologies, telles que le cancer ainsi que le guidage précis du geste chirurgical.

Deux types de rayonnement sont utilisés :

- Rayonnement monophotonique : radio-élément se désintégrant en émettant un rayonnement gamma sous forme d'un photon simple (scintigraphie gamma). L'image obtenue peut être bidimensionnelle (scintigraphie planaire) ou tridimensionnelle (tomographie d'émission monophotonique).
- Émission de positons Rayonnement biphotonique : chaque positon émis de nature instable, s'annihile avec un électron en produisant deux photons gamma de 511 KeV (tomographie par émission de positons ou TEP). L'image obtenue est toujours tridimensionnelle.

**A. Scintigraphie :** La scintigraphie est une technique d'imagerie qui explore la distribution d'un traceur radioactif dans un organe cible (thyroïde, os, cœur, etc.). Elle offre aux cliniciens la possibilité d'étudier la morphologie de cet organe et de détecter précocement les anomalies métaboliques fonctionnelles.

**B. Tomographie par Émission Monophotonique :** La Tomographie par Émission Monophotonique (TEMP), dite aussi SPECT (Single Photon Emission Computed Tomography), est une technique d'imagerie médicale nucléaire tomographique basée sur la scintigraphie. La machine de la SPECT permet de produire des images 3D via l'emploi de caméras gamma qui tournent autour du patient.

**C. Tomographie par Émission de Positons :** La TEP est une modalité d'imagerie conçue dans les années 60-70 mais dont l'introduction dans le domaine du diagnostic clinique est récente. Elle peut être associée au scanner on parlera alors d'équipement TEP-CT grâce au quel les images du scanner seront couplées à celles de la TEP. L'examen TEP permet de suivre la réponse du patient à un traitement

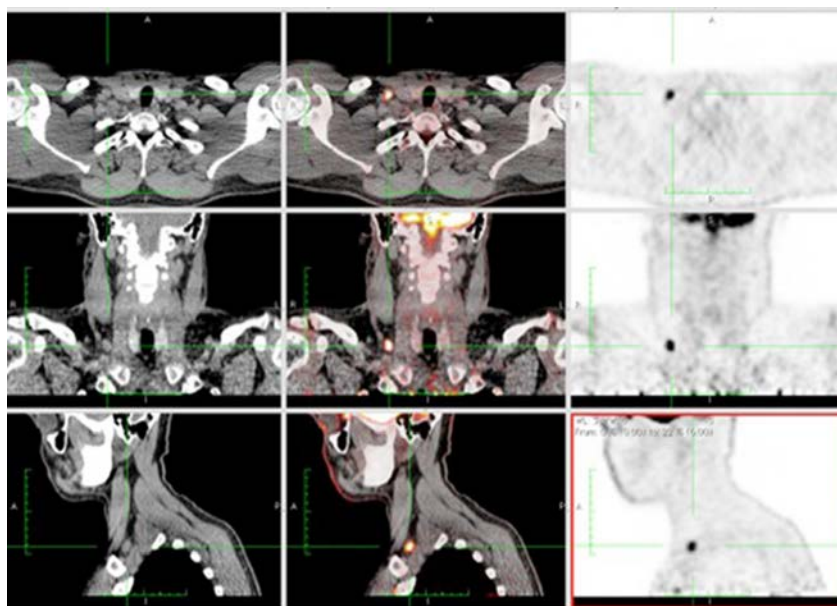


FIGURE 2.8 – Utilisation du traceur  $^{18}\text{F}$ -FDG pour surveiller la réponse thérapeutique en cancérologie. Cas clinique d'un patient avec Hodgkin Lymphome, évalué après huit séances de chimiothérapie : image CT (cadre à gauche), image TEP au  $^{18}\text{F}$ -FDG (cadre à droite) et des images de fusion TEP/CT (cadre au centre). La région de concentration du traceur FDG correspond à la localisation d'un petit ganglion supraclaviculaire droit sur l'image CT, indiquant clairement une récurrence lymphome Hodgkin, extraite de [ESR 2010].

de radiothérapie ou de chimiothérapie. La TEP permet la détection et la quantification des processus moléculaires tels que le métabolisme du glucose, angiogenèse, apoptose et nécrose [West 2014] [Rudin 2003].

En 1993, la TEP  $^{18}\text{F}$ -FDG a montré son efficacité dans la surveillance précoce de la réponse tumorale du cancer du sein [Wahl 1993], on note que le traceur FDG est utilisé pour l'évaluation de l'activité glycolytique du tissu tumoral. Depuis, il y a eu un intérêt croissant pour l'utilisation de la TEP  $^{18}\text{F}$ -FDG pour évaluer rapidement si une tumeur répond ou non au traitement dans plusieurs maladies [Larson 1999] [Weber 2006] (Figure 2.8). La TEP fournit aussi aux équipes médicales des diagnostics précoces, par exemple Di Chiro et al. ont démontré que la PET  $^{18}\text{F}$ -FDG pourrait aider à distinguer la nécrose tumorale cérébrale d'une tumeur active à la fin de la thérapie, même si la différence n'est pas claire sur des images CT [Di Chiro 1988a] [Di Chiro 1988b].

Pour résumer, l'emploi des techniques d'imagerie quantitative va principalement aider à : éliminer la subjectivité dans le processus d'analyse d'images, assurer la génération automatique des résultats d'imagerie, faire des corrélations statistiques entre les bases de données quantitatives et faciliter la corrélation entre les modèles de phénotypes d'imagerie et les profils génomiques. Ici, on s'est contenté de la présen-

tation des quelques techniques d'imagerie, la liste donnée n'est donc pas exhaustive.

## 2.4 L'ère des biomarqueurs d'imagerie

### 2.4.1 Définition du concept de biomarqueur d'imagerie

Au début des années 2000, l'autorité américaine de santé le « National Institute of Health » (NIH) a défini le concept de biomarqueur comme « une caractéristique qui est objectivement (c'est à dire avec une précision et reproductibilité suffisantes) mesurée et évaluée comme un indicateur de processus biologiques normaux, ou pathologiques ou de la réponse biologique à une intervention thérapeutique » [Downing 2001]. Cette définition du concept de biomarqueur a été adoptée par l'organisation américaine « Food and Drug Administration » (FDA)<sup>1</sup> [Katz 2004] et elle est maintenant largement reconnue dans la littérature scientifique. La dénomination « biomarqueur » est nouvelle mais le concept lui-même est employé en médecine clinique depuis des années ; par exemple la mesure de la glycémie est utilisée depuis l'année 1848 pour caractériser le diabète et évaluer l'efficacité des molécules antidiabétiques. Les biomarqueurs existent sous trois différentes formes :

- Paramètres biochimiques ou histologiques détectés à partir d'échantillons de tissus obtenus lors d'une biopsie ou d'une chirurgie,
- Paramètres biochimiques ou cellulaires obtenus à partir du sang ou d'échantillons d'urine,
- Paramètres anatomiques, fonctionnels ou moléculaires détectés par imagerie médicale.

De façon corollaire, un biomarqueur d'imagerie quantitative est défini comme un biomarqueur qui est dérivé à partir d'images médicales. Le groupe QIBA (Quantitative Imaging Biomarker Alliance) de la RSNA donne une définition stricte à ce concept, à savoir "une caractéristique objective dérivée à partir d'une image in vivo mesurée sur un rapport ou une échelle d'intervalle comme indicateur d'un processus biologique normal, d'un processus pathogène, ou une réponse à une intervention thérapeutique" [Kessler 2015]. Cette définition permet de supporter et guider la recherche en imagerie quantitative sur les biomarqueurs qui peuvent être évalués et comparés à l'aide de calculs statistiques (fréquences, moyennes, écarts types, et des erreurs standard de la moyenne). Le concept de biomarqueur d'imagerie a pris récemment une importance croissante, notamment dans le contexte du développement de la médecine personnalisée [Garcia-Donas 2013] et des médicaments [Abramson 2014] [O'Connor 2008].

Les biomarqueurs d'imagerie sont des paramètres qui estiment des propriétés anatomiques et physiologiques pertinentes comme la détection de la présence ou non d'une maladie, la prédiction de l'évolution de la pathologie (sans ou avec un traitement), l'évaluation de la réponse thérapeutique à un traitement [ESR 2010] [Prescott 2013].

---

1. <http://www.fda.gov/Drugs/>

Quelque soit le type de biomarqueur d'imagerie, un biomarqueur d'imagerie doit être quantifiable, reproductible (applicable à différentes populations de patients et par différents observateurs), précis (répétabilité), standardisé (protocoles définissant la méthode de mesure de acquisition à l'analyse des données) et qualifié pour différents usages dans des contextes différents (par exemple, corrélation avec la réponse thérapeutique reflétant la réponse positive ou négative du traitement).

### 2.4.2 Utilité des biomarqueurs d'imagerie

Les biomarqueurs d'imagerie jouent un rôle important dans la pratique clinique pour l'amélioration de la prise en charge des patients, ainsi que pour la recherche clinique pour le développement des nouveaux traitements notamment le développement des nouveaux médicaments [ESR 2010] [Buckler 2011b] [Brodbelt 2014] [Smith 2014] :

Au niveau de la délivrance des soins, le développement de la médecine de précision met en avant l'utilisation croissante de biomarqueurs quantitatifs dans tous les processus décisionnels, qu'ils soient en relation avec la prévention, le diagnostic, le choix thérapeutique et son suivi, en prenant notamment en compte les caractéristiques individuelles du patient dans l'estimation des facteurs pronostiques et prédictifs. Les biomarqueurs d'imagerie interviendront de façon significative dans ces futurs modèles décisionnels.

En recherche clinique, les biomarqueurs d'imagerie peuvent se substituer aux critères classiques de "clinical endpoints" ou "surrogate endpoints" ; ils sont susceptibles de permettre de tester de nouveaux médicaments à un stade asymptomatique précoce de la pathologie, ce qui pourrait conduire à une réduction de la durée des expérimentations, et donc une réduction significative des coûts.

#### 2.4.2.1 Pratique clinique : personnalisation des soins délivrés

Le développement des techniques d'imagerie quantitative incite de plus en plus à adopter des données quantitatives dans la pratique clinique [Boone 2007] [Choong 2012] et donc à changer la façon d'interpréter les images médicales. L'emploi des biomarqueurs d'imagerie a montré son efficacité dans différents domaines cliniques [Rosenkrantz 2015]. En cancérologie, plusieurs indicateurs d'imagerie dans le domaine du cancer sont employés en pratique clinique pour distinguer la tumeur de l'abcès bactérien [Hutter 2003], mesurer quantitativement l'efficacité de la réponse tumorale à une thérapie [Eisenhauer 2009], identifier les changements fonctionnels des tissus [Desar 2009] [Desar 2009], etc. De plus, les biomarqueurs d'imagerie sont utiles pour stratifier les patients et identifier d'une manière précoce ceux qui auront une mauvaise réponse à une thérapie donnée [Garcia-Donas 2013]. Ces indicateurs jouent un rôle croissant dans plusieurs domaines de la pathologie tels que les glioblastomes [Kao 2013], les maladies neurodégénératives, par exemple la maladie d'Alzheimer [Dickerson 2005], la sclérose en plaques [Miller 2004] [Filippi 2014], le cancer du sein [Wahl 2009] [Wahl 1993].



**A. Utilité clinique des biomarqueurs d'imagerie morphologiques :** La mesure de la taille d'une tumeur aide à évaluer l'évolution de la tumeur au cours du temps ; par exemple les *critères RECIST 1.1 (Response Evaluation Criteria in Solid Tumors)* qui sont basés sur la mesure du plus grand diamètre des lésions [Eisenhauer 2009] sont utilisés comme référence de l'évaluation de la réponse thérapeutique dans la prise en charge clinique des patients ainsi que les essais thérapeutiques. La méthode RECIST repose sur la définition de lésions cibles et de lésions non cibles, et la mesure des grands diamètres des lésions cibles. Le calcul du volume tumoral d'une tumeur et le calcul de la variation du volume tumoral ont été introduits pour fournir une estimation plus précise de la taille des lésions chez un patient [Moertel 1976] [Lavin 1980]. On peut distinguer deux types de variation du volume tumoral : une évolution positive appelée *Réponse* ou bien une évolution négative appelée *Progression*.

Une Réponse peut être une réponse complète : disparition de toutes les lésions ou une réponse partielle : diminution  $>30\%$  du plus grand diamètre des régions qui prennent le contraste de la lésion ou des lésions (la somme de diamètres), dans les deux cas il ne faut pas visualiser une progression des lésions non cibles. Une Progression est caractérisée par une augmentation d'au moins  $20\%$  de la somme des diamètres des régions qui prennent le contraste dans les nouvelles lésions, une augmentation indiscutable du volume des lésions non cible ou l'apparition de nouvelles lésions. Le compte rendu devrait comprendre un des termes suivants "CR" (Complete Response), "PR" (Partial Response), "SD" (Stable Disease) et "PD" (Progressive Disease). Le tableau 2.2 résume les réponses tumorales dans RECIST 1.1 et la Figure 2.9 représente un exemple d'évaluation de l'évolution d'une lésion ganglionnaire chez une patiente : l'augmentation de la taille des plus grands diamètres des lésions est supérieure à  $45\%$  (de 37.2 mm à 51.7mm), sans apparition de nouvelles lésions, donc la maladie est en progression.

Dans la méthode RECIST, les radiologues peuvent catégoriser les lésions observées sur les images CT ou IRM de la manière suivante :

- Lésion mesurable et cible : est une lésion choisie par l'opérateur pour faire les mesures, elle permet l'évaluation de la Réponse ou de la Progression.
- Lésion mesurable et non cible : est une lésion non choisie par l'opérateur (pour des raisons de taille, de nombre, ou de difficulté de mesure), elle permet l'évaluation de la Progression mais pas de la Réponse.
- Lésion non mesurable : est une lésion dont le diamètre est impossible à mesurer d'une manière précise, ce type de lésion permet d'évaluer la Progression mais pas la Réponse.

**B. Utilité clinique des biomarqueurs d'imagerie fonctionnelle :** Les biomarqueurs fonctionnels permettent l'estimation des processus biologiques comme la prolifération, l'hypoxie, l'angiogenèse, l'envahissement et la mort cellulaire [Hanahan 2000].

Premier exemple : La mesure *SUV (Standardized Uptake Value)* [Buvat 2007]

Évaluation \ Lésion	Cible	Non cible
Réponse complète	disparition	disparition
Réponse partielle	diminution d'au moins 30% de la somme des plus grands diamètres des lésions cibles	pas de progression
Maladie stable	ni réponse ni progression	pas des progression
Progression	augmentation supérieure à 20% de la somme des plus grands diamètres des lésions cibles	progression indiscutable de la taille des lésions ou d'une nouvelle lésion

TABLE 2.2 – Réponses tumorales dans RECIST 1.1

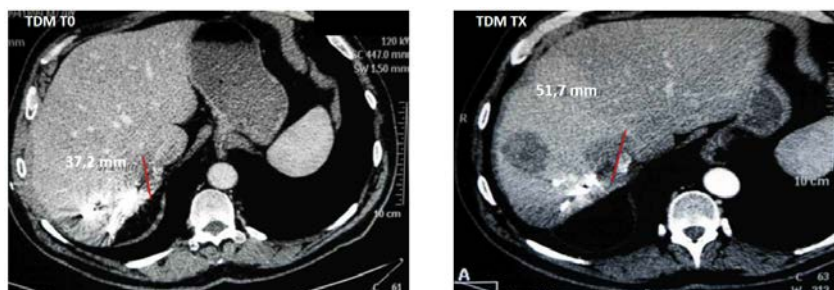


FIGURE 2.9 – Exemple d'évaluation de la progression d'une lésion ganglionnaire sur une TDM à quelques mois. Augmentation des valeurs des plus grands diamètres des lésions de 37,2mm à 51,7mm [Chan 2007].

quantifie la fixation tissulaire du traceur radioactif au sein du tissu tumoral dans une région d'intérêt. La valeur SUV se calcule comme le rapport de la fixation dans un tissu d'intérêt (en kBq/ml) à un instant donné à la dose injectée au patient (en kBq) rapportée à son volume (en ml) et elle est normalisée à la masse corporelle (kBq/g) du patient (Formule mathématique 2.1).

$$SUV = \frac{\text{concentration du traceur dans le tissu d'intérêt (KBq/mL)}}{\text{dose injectée à un temps zéro/poids corporel(g)}} \quad (2.1)$$

La valeur SUV est utilisée par les cliniciens pour caractériser l'activité d'une tumeur qui est essentielle pour la détection précoce des lésions tumorales malignes [Fletcher 2008] et de leurs récurrences, ainsi que pour le choix des traitements et l'évaluation de leur efficacité [Kinahan 2010] [la Fougère 2011]. Weber et al. [Weber 2009] montrent que les changements longitudinaux de la fixation du tra-

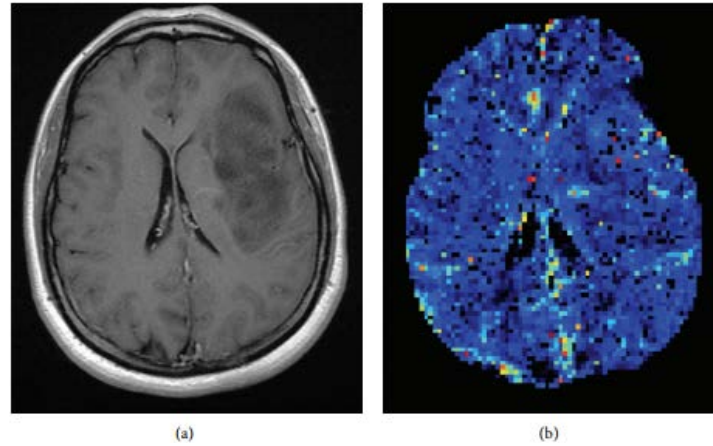


FIGURE 2.10 – (a) L'image pondérée T1 contrastée représente un astrocytome de classe II selon la classification de l'Organisation Mondiale de la Santé (OMS) (bas grade). (b) La carte  $K_{trans}$  montre aucune fuite du produit de contraste dans la région tumorale, ce qui suggère une faible valeur de  $K_{trans}$  [Kao 2013].

ceur  $^{18}F$ -FDG dans la région tumorale pendant les thérapies peuvent dans la majorité des cas prédire les résultats cliniques de façon plus précoce que les changements des mesures anatomiques. Au cours des 20 dernières années, l'utilisation du TEP au  $^{18}F$ -FDG dans le cancer a connu une croissance remarquable, les images de TEP étant de plus en plus utilisées de façon routinière dans le diagnostic, la détermination du stade de la maladie, et le suivi du traitement de nombreux cancers. Les critères PRECIST [Wahl 2009] permettent de faire l'évaluation de la réponse tumorale métabolique avec la valeur TEP au  $^{18}F$ -FDG.

Deuxième exemple : En imagerie de perfusion, le paramètre  $K_{trans}$  permet de mesurer un autre aspect de la tumeur qui est la perméabilité capillaire [Tofts 1999]. Le paramètre  $K_{trans}$  est calculé en mesurant le flux de fuite d'agent de contraste dans les espaces extravasculaires et extracellulaires. La valeur de  $K_{trans}$  permet d'évaluer la réponse thérapeutique d'un cancer de sein [Yankeelov 2007], de détermination du grade des gliomes (voir Figure 2.10) [Provenzale 2002], etc.

Troisième exemple : Le débit sanguin cérébral (DSC), le volume sanguin cérébral (VSC) et le temps de transit moyen (MTT) sont les principaux paramètres intervenant dans la perfusion cérébrale. La valeur du MTT indique le temps moyen nécessaire à la traversée des capillaires par le produit de contraste (en seconde), la valeur du VSC quantifie la quantité de parenchyme occupée par les vaisseaux sanguins (ml/100g), la valeur du débit sanguin cérébral est calculée en appliquant la formule 2.2.

$$DSC = \frac{VSC \text{ ml}/100g}{TTM \text{ min}} \quad (2.2)$$

Le volume cérébral régional est un bon indicateur de la densité micro-vasculaire dans le tissu tumoral, la mesure peut être employée pour établir le premier bilan de

diagnostic afin d'orienter sur la nature lésionnelle [Hu 2012] (maligne, s'il y'a une hyper-perfusion c'est à dire une néo-angiogenèse), faire le suivi post-thérapeutique [Barajas J.R. 2009], détection d'une transformation maligne d'une lésion de bas grade [Bisdas 2009], etc.

**D. Utilité clinique des biomarqueurs qualitatifs d'imagerie :** Les biomarqueurs d'imagerie qualitatifs (la définition du terme biomarqueur qualitatif est donnée dans la partie 2.4.5.3) servent à standardiser la description des observations radiologiques afin de minimiser la variabilité d'évaluation inter et intra observateurs. Ici, on citera deux terminologies :

- La terminologie mammographique BI-RADS (Breast Imaging Reporting and Data System) issue des travaux de l'"American College of Radiology" [Kopans 1992] définit des scores qualitatifs pour la description de l'aspect du sein en mettant en évidence quelques anomalies telles que les masses, calcifications, désorganisations architecturales, etc. Une fois ces anomalies détectées, les médecins se basent sur des critères qualitatifs pour caractériser la forme (rond, ovale, autre), l'orientation (parallèle, non parallèle), le contour (net/flou, fin/épais/mixte, lisse, régulier/irrégulier, spiculé), la texture (hétérogène/homogène/complexe), etc.
- La terminologie VASARI (Visually Accessible Rembrandt Images) [Archive 2013] a été proposée par le "Cancer Imaging Archive" (TCIA). Elle définit trente critères d'imagerie pour l'évaluation des tumeurs de type glioblastome. Les termes de cette terminologie décrivent la localisation de la lésion (hémisphère, lobe, etc.), la morphologie du bord de la lésion (épaisseur, prise de contraste, etc.), morphologie de la substance de la lésion (invasion des régions concernées, kyste, etc.), etc.

On note que ces deux terminologies contrairement aux critères RECIST prennent en considération les aspect de l'hétérogénéité des tumeurs solides en cancérologie. On peut citer d'autres terminologies : le Prostate Imaging Reporting and data system (PI-BRADS) [Bomers 2014] et le Lung Imaging and Reporting system (Lung BRADS) [Kazerooni 2016].

En imagerie médicale, des corrélations ont été démontrées entre des critères d'imagerie et des altérations génétiques liées à la pathologie [Diehn 2008] [Gutman 2013]. Dans [Diehn 2008], Diehn et al. ont montré l'existence d'une corrélation entre la prise de contraste et l'effet de masse et les gènes responsables de l'activation de la prolifération et l'hypoxie de la tumeur. Dans [Gutman 2013], Gutman et al. ont prouvé que la taille et la morphologie de la tumeur sont des critères prédictifs de la survie de malade.

#### 2.4.2.2 Recherche clinique ou translationnelle : développement des médicaments

Les biomarqueurs d'imagerie valides peuvent améliorer le développement de médicaments en facilitant, par exemple, la sélection des agents prometteurs en

vue des phases de recherche ultérieures ou bien la détermination des sous-groupes de patients dans lesquels la probabilité de réponse au médicament est plus élevée [Abramson 2014] [O'Connor 2008] [Frank 2003]. Le processus de recherche et développement des médicaments comporte trois grandes phases (Figure 2.11) [Garrett 1999] : la phase de recherche, la phase de développement pré-clinique et la phase de développement clinique. Les étapes de développement des différentes phases peuvent être décrites comme suit :

- **Phase de recherche** : étude théorique
- **Études précliniques** : administration d'une nouvelle substance ou entité biologique et l'évaluation de son efficacité et de sa toxicité sont évaluées chez les animaux.
- **Études cliniques** : les essais de l'étape de développement clinique sont des études effectuées chez l'Homme pour prouver la validité d'un nouveau traitement, démontrer l'existence d'un rapport bénéfice/risque pour les malades ciblés et de définir son schéma thérapeutique optimum. Les phases qui composent cette étape sont décrites ci-dessous :
  - Phase 0 : première administration de la substance chez l'Homme dans le but de déterminer le profil pharmacocinétique de la substance testée.
  - Phase 1 : définition du schéma thérapeutique (dose d'administration, fréquence d'administration, durée de traitement, etc.) et détermination de son degré de toxicité.
  - Phase 2 : test de l'efficacité biologique appelé aussi essais exploratoires qui consistent à réaliser un « test pilote » sur une population déterminée de patients en nombre réduit.
  - Phase 3 : efficacité clinique (par exemple la survie) appelé aussi essais confirmatoires, études statistiques sur des populations en nombre satisfaisant de patients qui consistent à évaluer l'effet thérapeutique (bénéfices) des médicaments sur des critères cliniques ainsi les éventuels risques (toxicité) du médicament.
  - Phase 4 : évaluation de l'efficacité après la commercialisation du médicament, évaluation de son effet thérapeutique et de sa toxicité sur des patients pris en charge.

Ainsi, d'après le FDA les biomarqueurs d'imagerie peuvent être utilisés pour différents intérêts (Figure 2.12) :

- Surveiller la sécurité d'une thérapie ;
- Déterminer si un traitement a l'effet désiré sur le corps ;
- Prédire les patients qui pourraient mieux répondre à un médicament à partir d'une perspective d'efficacité ;
- Permettre des économies de temps et de coûts dans les essais cliniques ;
- etc.

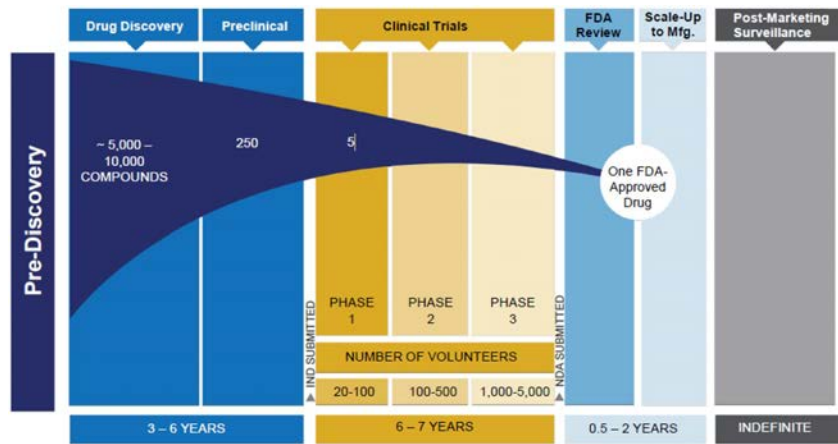


FIGURE 2.11 – Processus de recherche et de développement des médicaments [FDA 2005]

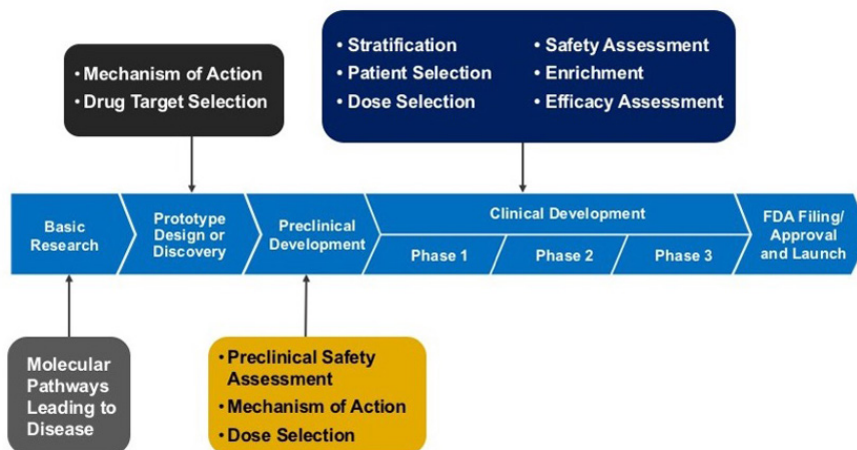


FIGURE 2.12 – Quelques exemples d'usage des biomarqueurs dans les différentes phases de R&D des médicaments, extraite de [FDA 2005]

### 2.4.3 Standardisation des biomarqueurs d'imagerie

La standardisation des résultats d'imagerie et de la structure des comptes rendus radiologiques est importante [Buckler 2011a] [Abramson 2014] [Sullivan 2015] pour assurer leur reproductibilité, améliorer la communication entre les équipes de recherche (au sein et entre différents centres de recherche qui travaillent sur la même étude) et entre les médecins spécialistes et les radiologues (au sein de l'hôpital ou bien entre les hôpitaux). Pour cela, le groupe QIBA de la RSNA [RSNA 2007] a créé des profils de standardisation qui ont comme objectif d'explicitier d'une manière détaillée la définition, la génération et l'utilisation de ces indicateurs pour pouvoir faciliter leur réutilisation (Figure 2.13 représente la structure générale du profil QIBA). Ces profils détaillent les dépendances entre les biomarqueurs d'imagerie et les protocoles d'acquisition d'images, on peut citer quelques exemples de profils QIBA actuellement disponibles :

- Le profil QIBA "CT tumor change volume" : Ce profil propose des méthodes standardisées pour mesurer le volume d'une lésion d'une manière reproductible pour évaluer la réponse thérapeutique des patients. Ce profil considère qu'il y a un changement réel dans le volume tumoral (avec un niveau de confiance égal à 95%) que si la variation mesurée est supérieure à 24%, 29% ou 39%.
- Le profil QIBA "FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy" : Les spécifications techniques décrites dans ce profil sont appropriées à la quantification de l'absorption du traceur FDG par les tissus tumoraux (la valeur  $SUV_{max}$ ) et la mesure des changements longitudinaux chez les sujets. Des approches normalisées de la mesure de la valeur du PET et des processus d'étalonnage de machines sont décrites dans le document. A titre d'information le FDG est un traceur qui permet d'étudier l'augmentation de la consommation du glucose par les tissus tumoraux. Une augmentation de la valeur  $SUV_{max}$  de 39% ou plus, ou une diminution de -28% ou plus indique qu'un véritable changement s'est produit avec une confiance de 95%. Une description détaillée de ce profil est donnée dans la sous-section 2.4.3.1.
- Le profil QIBA "DWI-MRI for tumor tissue characterization" : Ce profil spécifie les différentes étapes de mesure du coefficient ADC. Un changement de 26% ou plus de la valeur ADC d'une lésion hépatique indique qu'un véritable changement a eu lieu avec 95% de confiance.
- etc.

Ces profils QIBA sont le résultat de coopérations entre différents experts en développement, analyse et examen réglementaire des biomarqueurs, etc. Les méthodes définies couvrent les différentes étapes du processus de mesurage du biomarqueur : la préparation de patient, l'acquisition d'image, le traitement d'image, l'analyse d'image et l'archivage des valeurs du biomarqueur. Les "futurs" utilisateurs de ces profils QIBA, tel qu'ils le précisent, sont principalement : les fabricants de logiciels et d'appareils qui créent des produits de mesurage des phénotypes d'imagerie, les

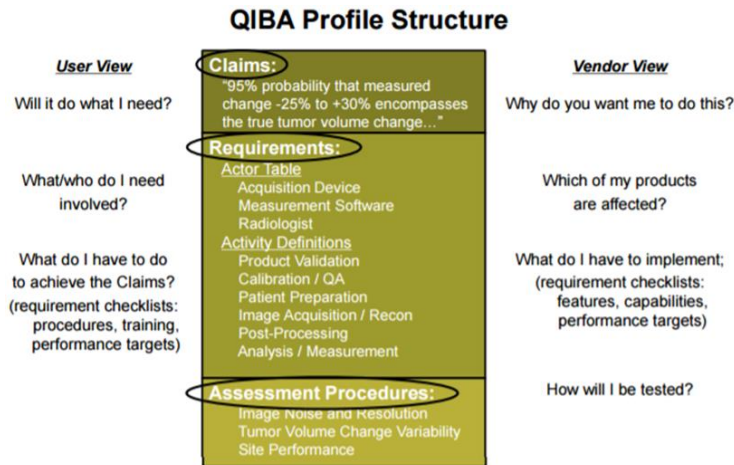


FIGURE 2.13 – Structure standard d'un profil QIBA [Kevin 2016].

sociétés biopharmaceutiques, les oncologues, les scientifiques des essais cliniques qui conçoivent des essais avec des paramètres d'imagerie, les radiologues, les médecins spécialistes, etc.

#### 2.4.3.1 Le profil QIBA "FDG-PET/CT as an imaging biomarker measuring response to cancer therapy"

Le paramètre SUV a différentes variantes qui sont le  $SUV_{max}$ ,  $SUV_{mean}$ ,  $SUV_{peak}$  ou RSUV ;  $SUV_{max}$  et  $SUV_{mean}$  sont des valeurs absolues alors que RSUV est un rapport entre des valeurs de SUV concernant différentes régions anatomiques. Ici, le terme  $SUV_x$  sera utilisé pour faire référence à toutes ces différentes possibilités de mesure. Les multiples façons de quantifier la valeur  $SUV_x$  rend difficile sa réutilisation dans les études. Ce profil QIBA aborde le problème de la quantification de la valeur  $SUV_x$  avec le traceur FDG et du mesurage de sa variation afin de mieux évaluer la réponse thérapeutique en cancérologie, le changement de la consommation du traceur FDG est calculé avec la formule 2.3 ;  $Ametab_{pré}$  désigne l'activité métabolique avant le traitement et  $Ametab_{post}$  désigne l'activité métabolique après le traitement.

$$\frac{Ametab_{post} - Ametab_{pré}}{Ametab_{pré}} \times 100 \quad (2.3)$$

Ici, on listera quelques contraintes à prendre en compte selon le profil QIBA (les étapes de mesurage de la valeur  $SUV_x$  sont décrites dans la Figure 2.14) :

1. **Préparation et positionnement du patient** : avant l'examen, le sujet doit être à jeun au moins 6 heures avant l'injection du produit radiopharmaceutique afin de réduire la consommation du glucose par les tissus sains. La standardisation de l'étape de préparation du patient entre les sites est importante pour éviter la distribution hétérogène du traceur FDG.



2. **Préparation et administration du radio-traceur** : le poids et la taille du patient, la quantité de dose injectée, la quantité de radio-traceur non injectée dans le corps du patient, le temps d'injection, le niveau de glucose dans le sang, etc. Tous ces paramètres doivent être enregistrés pour réaliser de futures explorations du même patient dans des conditions similaires et rendre possibles les comparaisons des valeurs de SUV.
3. **Acquisition des images TEP/CT** : cette étape ne peut être réalisée qu'au delà d'une période d'au moins 60 min après l'injection du produit radio-pharmaceutique, condition nécessaire pour assurer que l'instrument d'acquisition d'image va donner les mêmes résultats au cours de l'étude. Pour cette même raison, les paramètres d'acquisition des appareils TEP (temps par position de lit, mode d'acquisition 2D ou 3D) et CT (temps, kvp, pitch, épaisseur de la coupe) doivent être enregistrés.
4. **Traitement des images TEP et CT** : les images CT sont reconstruites en coupes axiales et réorientées en coupes sagittales ou coronales, pour faciliter la localisation des anomalies scintigraphiques détectées sur les images TEP. Une correction de l'atténuation est généralement réalisée, si nécessaire aux images TEP.
5. **Mesure quantitative de la valeur  $SUV_x$**  : la quantification de la mesure  $SUV_x$  telle qu'elle est définie antérieurement est un index qui mesure la captation par la tumeur, normalisée par l'activité injectée et le poids du patient. Le choix de la méthode de calcul du biomarqueur d'imagerie (calcul de la valeur moyenne, calcul de la valeur maximale ou calcul de la valeur Pic) doit être indiqué.
6. **Interprétation qualitative des images TEP/CT** : comparaison entre les données quantitatives obtenues et les données visuelles. Plusieurs manipulations (zoom, interpolation bi-linéaire/ tri-linéaire, ré-orientation, etc.) peuvent être faites à ce niveau pour améliorer l'affichage des images médicales ou bien pour normaliser le calcul de la valeur SUV au sein du tissu.

#### 2.4.4 Processus de développement des biomarqueurs d'imagerie

Le processus de développement des biomarqueurs d'imagerie repose sur la validation de l'existence d'une relation entre le biomarqueur en question et la réalité objective étudiée (physiologique, biologique, morphologique ou moléculaire). Les étapes de développement des biomarqueurs d'imagerie sont définies dans [Ball 2010] [Bonmatí 2012] [ESR 2013], la figure 2.15 illustre ces différentes étapes :

- **Preuve de la faisabilité issue de la recherche académique** : consiste en la définition de l'objectif de l'étude de l'aspect biologique de la maladie exploré par le biomarqueur.
- **Preuve du mécanisme** : cette étude est réalisée pour la détermination de la nature (proportionnelle, inversement proportionnelle, linéaire, etc.) de la

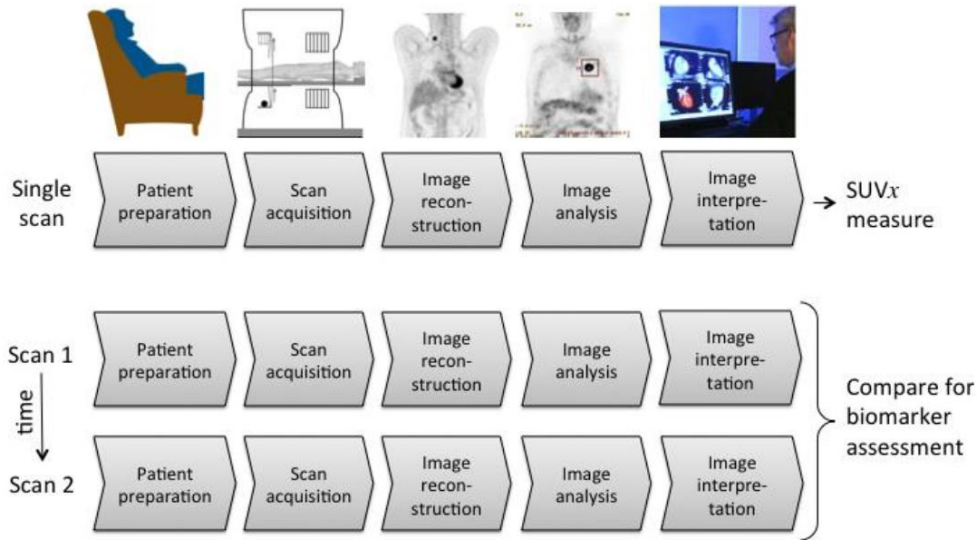


FIGURE 2.14 – Les étapes de mesure de la valeur SUV [RSNA 2007]

relation qui est établie entre le biomarqueur, l'aspect analysé et la maladie en cours d'étude.

- **Acquisition des images médicales** : des méthodes d'acquisition des images médicales sont définies pour assurer la mesure de la valeur du biomarqueur d'imagerie d'une manière reproductible.
- **Prétraitement des images médicales** : la qualité des images produites est améliorée via des méthodes d'imagerie afin de faciliter leur analyse. Souvent, deux paramètres doivent être optimisés pour améliorer le calcul de valeurs avec le moindre biais : rapport signal/bruit (SNR) et le rapport le contraste de bruit (CNR). Cette étape fait donc intervenir la réduction du bruit, le recalage des images et d'autres traitements d'images.
  - Réduction du bruit : l'étape d'acquisition peut mettre en jeu la correction des altérations avant/après l'analyse des images. Dans cette étape, on agit sur la variation de la luminosité de l'image qui peut être réduite par les filtres.
  - Recalage d'images : consiste à estimer une transformation géométrique entre les points de plusieurs images qui appartiennent à des espaces différents. Cette transformation permet de passer d'une image à une autre. Le calcul du volume cérébral par exemple nécessite l'application du recalage d'images. Plusieurs méthodes de recalage existent dans la littérature et leur classification est basée sur plusieurs critères : la dimension de l'image (2D ou 3D) et leurs modalités, la nature de la transformation (rigide, affine, non linéaire) et la procédure d'optimisation.
- **Analyse des images médicales et modélisation des paramètres**

**d'imagerie** : des méthodes d'extraction des informations relatives aux biomarqueurs d'imagerie sont définies afin de standardiser les techniques qui concernent par exemple, la délimitation de la région d'intérêt, le calcul du volume lésionnel et la sélection des modèles mathématiques d'étude des propriétés morphologiques, biologiques, fonctionnelles ou moléculaires [Zussman 2011]. Ces paramètres influencent les résultats et la reproductibilité des mesures, par exemple en imagerie fonctionnelle, la sélection des ROI a un impact sur le calcul de la valeur de SUV [Buckler 2011a]. Les valeurs extraites des images médicales sont représentées sur des cartes paramétriques (par exemple les cartes d'ADC) dont chaque voxel représente une valeur de biomarqueur d'imagerie. Une analyse visuelle des images médicale peut ainsi être faite à partir des modèles mathématiques de distribution de valeurs de biomarqueurs ou bien sur des cartes appelées multi-paramétriques qui représentent la distribution conjointe de plusieurs variables ; combinaison de différentes valeurs de biomarqueurs (analyse multi-variée).

- **Mesure de la valeur de biomarqueur d'imagerie** : les modèles mathématiques de calcul de la valeur de biomarqueur d'imagerie doivent être précisés. Dans la plupart des cas, des méthodes statistiques sont appliquées pour la représentation des valeurs mesurées sous forme d'histogramme mais d'autres façons peuvent être employés comme le calcul de Kurtosis qui est conseillé pour éviter les problèmes de sous estimation qui sont dus à l'hétérogénéité de la distribution des valeurs de biomarqueurs dans la région d'intérêt.

Une fois le biomarqueur d'imagerie découvert et vérifié (voir la partie 2.4.5.1 sur les différents états des biomarqueurs), sa performance doit être évaluée par des tests statistiques. Les principales étapes de cette phase d'évaluation d'après l' "Institute of Medicine" [Ball 2010], sont la validation, la qualification et l'utilisation (Figure 2.16).

- **Validation du biomarqueur d'imagerie** : la validation d'un biomarqueur d'imagerie est le processus d'évaluation de la performance de la méthode de mesure et la détermination des conditions de répétabilité (mesures successives effectuées dans les mêmes conditions) et de reproductibilité (accord entre des mesures successives effectuées avec des conditions variables telles que la localisation ou l'opérateur) sous lesquelles le biomarqueur fournira des valeurs exactes. En d'autres termes, l'étape de validation consiste à démontrer la faisabilité et la précision de la méthode de mesure des biomarqueurs.
- **Qualification du biomarqueur d'imagerie** : c'est l'étape de la validation technique de la méthode de mesurage du biomarqueur appelée aussi la preuve du principe, évaluation des performances et de la capacité à refléter la réalité biologique, (par exemple, processus pathologique ou la réponse à un traitement) du biomarqueur dans le contexte spécifique de son utilisation proposée [Wintermark 2008] [Bruns 2009] [Huynh 2014]. Cette étape est nécessaire dans l'évaluation du biomarqueur d'imagerie pour son adoption dans les applications cliniques et de recherche, les résultats de la phase de quali-

fication sont gardées par les autorités réglementaires comme des preuves de l'efficacité du médicament et de l'outil de mesurage [Katz 2004]. L'un des défis statistiques de la qualification est de démontrer qu'un biomarqueur est associé à un "clinical endpoint" par exemple (l'effet du traitement sur le biomarqueur d'imagerie doit corrélé étroitement avec l'effet du traitement sur le résultat clinique). En fait, un nombre très réduit de biomarqueurs d'imagerie répond aux critères de substitution (la notion de critère de substitution est décrite dans le paragraphe 2.4.5.2).

- **Utilisation du biomarqueur d'imagerie** : Ici ce sont des tests de preuve d'efficacité et d'efficience qui sont réalisés pour contrôler la qualité de la méthode de dosage lors de son utilisation dans la routine clinique [Green 2008] et ainsi les avantages et les inconvénients de l'utilisation du biomarqueur d'imagerie dans le contexte clinique.

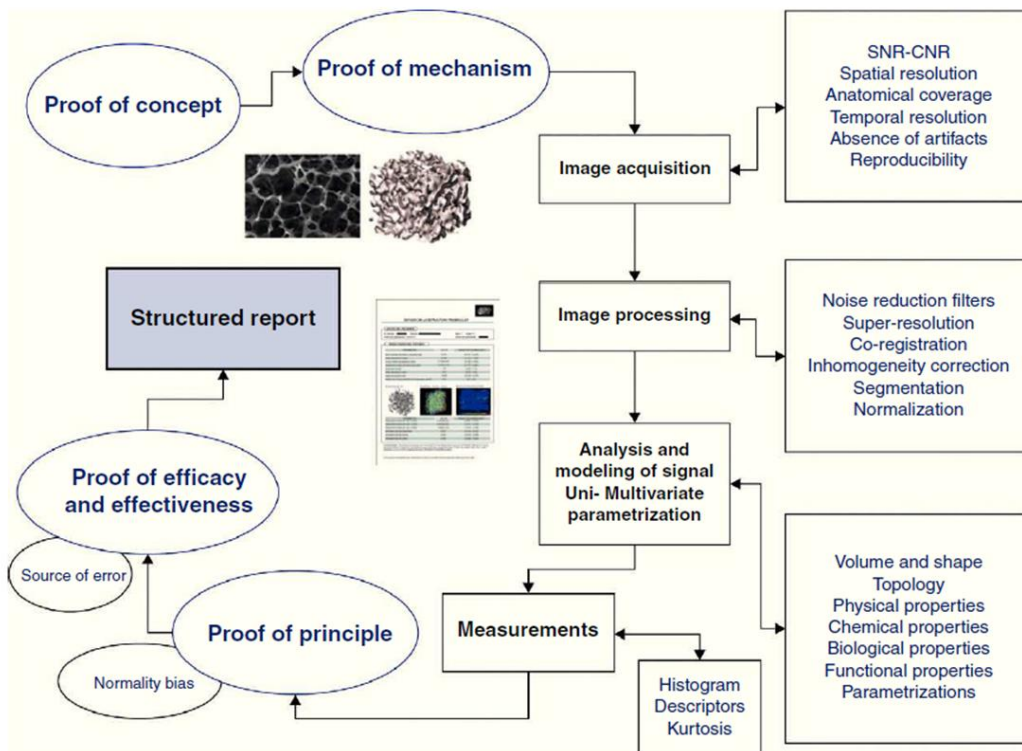


FIGURE 2.15 – Figure extraite de [ESR 2013] décrivant les principales étapes de développement d'un biomarqueur d'imagerie.

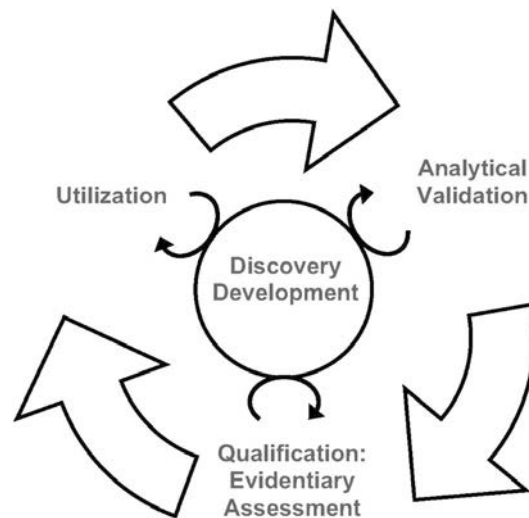


FIGURE 2.16 – Les étapes d'évaluation des biomarqueurs d'imagerie d'après le "Institute of Medicine" [Downing 2001].

## 2.4.5 Classification des biomarqueurs d'imagerie

### 2.4.5.1 Classification des biomarqueurs d'imagerie selon l'état de validation

Avant leur inclusion dans la routine clinique, les biomarqueurs d'imagerie passent par quatre « statuts réglementaires » de validation [Marrer 2007] : d'abord, ils sont *découverts* lorsque l'étude du mécanisme biologique lié à la maladie est encore à prouver. Ensuite, ils sont *vérifiés*, lorsque des démonstrations du mécanisme ont été réalisées par différents laboratoires. Puis ils sont *validés* ; d'après le FDA un biomarqueur est dit *validé* quand il est "mesuré par une méthode analytique aux performances établies et pour lequel il y'a suffisamment d'éléments scientifiques pour démontrer la signification clinique, toxicologique ou pharmacologique des données générées par le biomarqueur". Enfin, ils sont *qualifiés* lorsque les autorités de santé valident la pertinence clinique du biomarqueur après l'évaluation des données de validation fournies par le promoteur.

### 2.4.5.2 Classification des biomarqueurs d'imagerie selon leur rôle

Les biomarqueurs d'imagerie peuvent jouer différents rôles dans les processus décisionnels [ESR 2011] [Waterton 2008] [Buckler 2011c] :

- **Biomarqueur diagnostique d'imagerie** : est un biomarqueur d'imagerie qui est utile pour caractériser la maladie au plan biologique et donc classifier les individus.
- **Biomarqueur pronostique d'imagerie** : est un biomarqueur d'imagerie qui permet de différencier des catégories de patients présentant différents risques pour une évolution déterminée, indépendamment du choix du trai-

tement administré (ou du choix de ne pas administrer de traitement) ; il s'agit d'une prédiction du devenir de la maladie en l'absence de traitement (prédiction de l'issue spontanée).

- **Biomarqueur prédictif d'imagerie** : est un biomarqueur d'imagerie qui est destiné de prévoir la ou les éventuels bénéfices (efficacité) et risques (toxicité) d'un traitement s'il est administré
- **Biomarqueur de surveillance d'imagerie** : est un biomarqueur d'imagerie qui aide à prédire le résultat clinique final d'un patient particulier après qu'un médicament particulier (ou une autre intervention) ait été administré.
- **Biomarqueur de réponse** : est un biomarqueur d'imagerie qui aide à prédire l'efficacité d'un médicament et sa capacité de modifier le résultat clinique chez un groupe de patients dans les essais cliniques.
- **Critère d'évaluation clinique « clinical endpoint »** : est une caractéristique ou variable qui reflète l'état du patient. L'autorité de santé américaine a qualifié les biomarqueurs d'imagerie précédemment cités comme des biomarqueurs d'imagerie « simples » et elle a introduit la notion de « critère de substitution ».
- **Biomarqueur d'imagerie de substitution** : est appelé aussi « surrogate endpoint » pour qu'un biomarqueur d'imagerie puisse être classé comme un « critère de substitution » il doit permettre d'étudier la relation entre l'état du patient et son devenir. Cette catégorie de biomarqueurs destinés à se substituer à un critère d'évaluation clinique devant permettre de déterminer le bénéfice clinique ou le risque à partir de données épidémiologiques, thérapeutiques ou physiopathologiques. De nombreuses données issues de plusieurs essais cliniques contrôlés réalisés avec des populations en nombre satisfaisant sont nécessaires afin d'établir la corrélation entre le biomarqueur concerné et la variation du paramètre biologique cible.

### 2.4.5.3 Classification des biomarqueurs selon leur type

Il est possible de classer les biomarqueurs d'imagerie selon la nature de la qualité mesurée : biomarqueurs quantitatifs d'imagerie et biomarqueurs qualitatifs d'imagerie.

Le groupe QIBA restreint la définition des biomarqueurs quantitatifs d'imagerie aux seules variables de type ratio ou d'intervalle. Cela implique que les volumes de tumeurs ou les valeurs de SUV par exemple, sont des biomarqueurs d'imagerie quantitatifs car la différence ou le rapport entre deux valeurs de ces mesures est "significative" [Kessler 2015].

**A. Biomarqueur quantitatif d'imagerie** Un biomarqueur d'imagerie quantitatif tel qu'il est défini par le groupe QIBA, est une caractéristique numérique qui est extraite à partir des techniques d'imagerie quantitative et qui a des propriétés de mesures quantitatives. Ici, nous présentons trois notions de base intrinsèques aux mesures quantitatives qui sont formellement définies

dans la troisième édition du Vocabulaire International de Métrologie (VIM) [Joint Committee for Guides in Metrology 2012] :

- Grandeur [VIM, 1.1] est une « propriété d'un phénomène, d'un corps ou d'une substance, que l'on peut exprimer quantitativement sous forme d'un nombre et d'une référence ; la référence peut être une unité de mesure, une procédure de mesure, un matériau de référence, ou une de leurs combinaisons ». Exemples de grandeurs génériques : longueur, énergie, charge électrique, etc.
- Nature de grandeur [VIM, 1.2] est un « aspect commun à des grandeurs mutuellement comparables ». Par exemple, les grandeurs diamètre, circonférence et longueur d'onde sont des grandeurs de même nature, à savoir la nature de la longueur.
- Valeur de la grandeur [VIM, 1.19] est un « ensemble d'un nombre et d'une référence constituant l'expression quantitative d'une grandeur ». La valeur mesurée d'un biomarqueur d'imagerie est le résultat de l'exécution du processus de mesurage. La valeur vraie est la valeur réelle qu'on cherche à estimer via le processus de mesurage. La valeur d'une grandeur peut être : un produit d'un nombre et d'une unité de mesure par exemple (5cm la taille d'une lésion) une grandeur sans dimension (la mesure du SUV), un nombre et la référence à une procédure de mesure, un nombre et un matériau de référence, etc. la Figure 2.17 représente les notions de base autour du concept de grandeur telles que définis dans le VIM.

**B. Biomarqueur qualitatif d'imagerie** D'après le VIM, une propriété qualitative (le concept 1.30 sur la Figure 2.17) est « une propriété d'un phénomène, d'un corps ou d'une substance, que l'on ne peut pas exprimer quantitativement », par exemple : l'absence ou la présence d'un kyste (oui, non), la qualité du contraste pris par la tumeur (aucune, moyenne, forte), etc. Ces valeurs peuvent être ordinales c'est à dire associées à une échelle de mesure qui ordonnent les différentes valeurs, ou non ordinales.

## **2.5 Besoin d'intégration des données d'imagerie pour supporter le développement des bio-banques d'images**

Ces vingt dernières années, le volume des données d'imagerie quantitatives décrivant le contenu des images médicales produites a vu une croissance exponentielle [DiSantis 1986] [Buckler 2013b]. La mise à disposition des médecins, chercheurs, biologistes, scientifiques, développeurs des technologies, etc. des infrastructures informatiques pour stocker des données d'imagerie sous forme numérique est devenu inévitable. Ces entrepôts de données sont appelés les bio-banques d'images et ils ont comme objectif de faciliter la diffusion des données d'une manière compréhensible

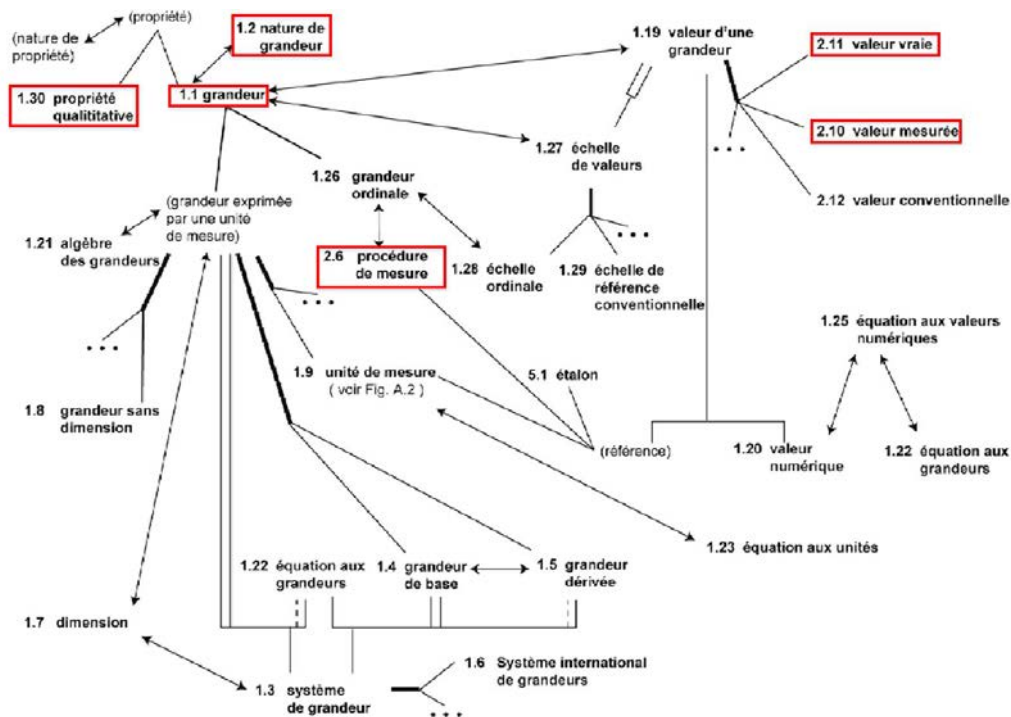


FIGURE 2.17 – Schéma conceptuel autour du concept grandeur. Les concepts encadrés sont les quelques notions clés qui sont abordées dans ce rapport. Pour mieux comprendre les différentes significations des relations de spécialisation entre les concepts il faut retourner à la source [Joint Committee for Guides in Metrology 2012].

pour correctement les exploiter (comparaison des traitements d'images, normalisation et validation, etc.).

Les bio-banques d'images [Hewitt 2013] sont définies en 2011 comme "des bases d'images médicales et de biomarqueurs d'imagerie associés (radiologie et au-delà), partagées entre plusieurs chercheurs et bases biologiques". Les bio-banques d'images sont donc des bases de données partagées de biomarqueurs d'imagerie qui sont liées à des banques d'échantillons biologiques et qui sont conçues pour donner un accès aux données qui concernent les images médicales ainsi que des informations sur les génotypes des sujets/patients. Ces bases de recherche vont guider les médecins dans la pratique clinique [Asslaber 2007] [ESR 2011] à réaliser des rapprochements entre les données génétiques des patients et les données d'imagerie pour faire des nouvelles découvertes des "génotypes" qui décrivent le mieux la pathologie, faciliter les études pluridisciplinaires, accélérer les développements de thérapies ciblées, etc.

Les premiers projets européens de bio-banques d'images sont l'UK Bio-Bank [Petersen 2013] et la cohorte nationale allemande [Woodbridge 2013] [Starkbaum 2014]. D'après [ESR 2015], il existe trois types de bio-banques d'images :



- les bio-banques d'images pour la recherche clinique qui permettent de archiver, de partage et de diffuser (pour usage secondaire) les données d'imagerie émanant des essais cliniques.
- les bio-banques d'images axées sur l'étude des pathologies qui étudient des patients qui sont atteints d'une certaine pathologie.
- les bio-banques d'images axées sur l'étude des populations qui étudient des sujets sains/malades.

L'intégration des données d'imagerie n'est pas une tâche facile, vu la variabilité des paramètres d'acquisition, des procédures de traitement et des méthodes de quantification des paramètres d'imagerie. Les données d'imagerie couvrent à la fois les images acquises, les images dérivées et les méta-données d'imagerie :

- **Images acquises** : les images brutes de tous types de modalité d'imagerie comme l'IRM, le CT, la TEP, etc.
- **Images dérivées** : les images issues des traitements d'imagerie par exemple de type dé-bruitage (e.g. images débruitées), segmentation (e.g. régions d'intérêt), recalage (e.g. images recalées), etc.
- **Méta-données d'images** : les informations décrivant les résultats de traitements (données dérivées des images médicales telles que les valeurs de biomarqueurs d'imagerie, les matrices GLCM (Gray-Level Co-occurrence Matrix), etc.), des informations de provenance (informations sur le traitement des informations outils de traitement, entrées, sorties et procédures), des observations faites par les radiologues, etc.

Les images médicales sont généralement enregistrées sous le format de standardisation DICOM<sup>2</sup>, des systèmes de gestion de dépôts d'images comme NBIA<sup>3</sup> et XNAT<sup>4</sup> peuvent être employés pour la manipulation des données. Les méta-données d'images peuvent être répertoriées dans des rapports structurés tels que : le DICOM SR [Clunie 2007a] dans le contexte clinique, le modèle AIM [Channin 2010] ou le eCRF (electronic Case Report Form) dans le cadre de la recherche. Cette diversité des outils et des plates-formes d'analyse utilisés dans l'imagerie quantitative est un frein majeur vis à vis du partage et réutilisation à grande échelle des données d'imagerie tel que l'a souligné le réseau national de recherche quantitative de l'institut national américain du cancer QIN (Quantitative Imaging Network) [Levy 2012]. Dans ce travail, le QIN a présenté le résultat d'un recensement des principales normes et outils utilisés par ses sites de recherche (le Tableau 2.3 reprend une partie du résultat présenté) et par la suite il a proposé une architecture générale du système d'information à mettre en place pour favoriser la collaboration dans les essais cliniques en cancérologie.

---

2. <http://medical.nema.org/>

3. <https://imaging.nci.nih.gov>

4. <http://www.xnat.org/>

<b>Modalités d'image</b>	CT de perfusion, IRM dynamique avec produit de contraste, IRM de perfusion, spectroscopie à résonance magnétique nucléaire, IRM d'imagerie de sodium, etc.
<b>Critères d'imagerie</b>	la longueur, le volume, la valeur du SUV, paramètres cinétiques de perfusion, l'hypoxie tumorale, etc.
<b>Outils d'acquisition d'images</b>	ePAD [Rubin 2008b], 3D Slicer <sup>5, 6</sup> , Matlab <sup>7</sup> , Definiens <sup>8</sup> , MimVista <sup>9</sup> , etc.

TABLE 2.3 – Principales données d'imagerie en cancer d'après les statistiques du QIN [Levy 2012]

## 2.6 Contexte d'étude : représentation des observations d'imagerie médicale

Rappelons que le domaine étudié dans le cadre de la thèse est celui du partage des données d'imagerie qui concernent la production et l'utilisation des paramètres quantitatifs et qualitatifs en imagerie dans les deux environnements de la délivrance des soins et de la recherche biomédicale, clinique et translationnelle. Dans cette section, on propose de lister quelques besoins fonctionnels du domaine.

### 2.6.1 Description des besoins médicaux dans les applications dédiées à la délivrance de soins

Les besoins fonctionnels concernent trois types d'activités ; des actions réalisées en amont de la production des biomarqueurs d'imagerie, des actions exécutées pour la production des valeurs des biomarqueurs d'imagerie et d'autres qui concernent l'utilisation de la valeur des biomarqueurs pour la prise de décision.

1. **Planification des actions réalisés en amont de la production de valeur des biomarqueurs d'imagerie** : Cette étape concerne la planification des tâches pour répondre au mieux à la question clinique à laquelle le médecin cherche à répondre. Elle inclut l'identification du ou des biomarqueurs d'imagerie adaptés au patient et la planification des étapes de traitement d'images à appliquer pour la production de la mesure de biomarqueur sélectionné.
2. **Exécution des actions pour la production des valeurs des biomarqueurs d'imagerie** : Cette étape concerne la réalisation des étapes d'imagerie décrites dans l'étape précédente : acquisition, reconstruction et analyse d'images afin de calculer la mesure.
3. **Utilisation des valeurs des biomarqueurs d'imagerie dans des processus décisionnels** : Cette étape consiste en l'application des biomarqueurs d'imagerie dans des processus décisionnels pour l'aide la décision.

**2.6.1.1 Cadre 1 : besoin de planification des actions avant la production des valeurs de biomarqueurs d'imagerie**

Le choix du biomarqueur d'imagerie le plus adapté au patient repose sur la nature de la question clinique (diagnostique, pronostique, d'évaluation de réponse à un traitement, etc.) et le contexte clinique. La relation de dépendance entre le contexte clinique, la procédure d'imagerie et la valeur du biomarqueur d'imagerie doit être formalisée pour automatiser l'identification et la planification des tâches de traitement nécessaires pour l'extraction de la valeur du biomarqueur d'imagerie. Les protocoles d'imagerie à appliquer pour extraire la valeur du biomarqueur doivent être déterminés par le type de biomarqueur à produire pour être ensuite décrits au médecin ; ces protocoles concernent l'acquisition, le pré-traitement et le traitement des images médicales. L'objectif de cette étape est de déterminer le ou les biomarqueurs d'imagerie qui correspondent le mieux au contexte clinique en question et fournir le plan d'action. Les données d'entrée sont alors les informations cliniques décrivant l'état du patient et la question clinique posée par le médecin. Les données de sortie sont le ou les biomarqueurs à mesurer, les modalités d'images à acquérir et les traitements et pré-traitements à appliquer. L'établissement de liens entre (1) la procédure de mesure (protocoles et procédures d'imagerie), (2) les critères cliniques intervenant dans la décision, (3) le/les rôles des biomarqueurs d'imagerie et (4) les modèles de décision est nécessaire.

L'évaluation de la réponse à un traitement en neuro-oncologie par exemple, peut se faire avec le calcul de la variation du volume lésionnel, le changement de la valeur SUV, les critères RECIST, etc. Supposons que le médecin a choisi la valeur SUV comme un critère d'évaluation. Dans ce cas, il doit suivre les protocoles du processus de calcul de la valeur SUV (les étapes du plan d'action du calcul de la valeur SUV sont décrites dans la partie 2.4.3.1).

**2.6.1.2 Cadre 2 : besoin d'exécution des actions pour la production des valeurs des biomarqueurs d'imagerie**

Cette partie concerne l'exécution des étapes spécifiées dans le plan d'action des processus de traitement d'images (acquisition et reconstruction) et d'extraction des valeurs de biomarqueurs d'imagerie. Cette étape consiste à traiter les images selon le protocole spécifié dans le plan d'action. Ceci peut comprendre des actions réalisées comme l'amélioration de la qualité d'image, la délimitation d'une région d'intérêt, la segmentation d'une tumeur, etc. La difficulté rencontrée au niveau de cette partie est que les traitements d'images et de calcul des biomarqueurs d'imagerie ne sont pas encore automatisés et que les données de provenance associées aux valeurs de biomarqueurs d'imagerie sont enregistrées selon des formats standards ou non. L'établissement de lien entre les processus de traitements et les protocoles va automatiser l'acquisition et le traitement des images avec l'équipement et selon le protocole spécifié dans le plan d'action.

### 2.6.1.3 Cadre 3 : besoin d'utilisation des valeurs des biomarqueurs d'imagerie pour l'aide dans la prise de décision

#### A. Implémentation des schémas thérapeutiques/ flux de travail basés sur les biomarqueurs d'imagerie

Le but de cette étape est de lier la question clinique avec les modèles de décision potentiellement applicables. Dans la pratique clinique, les biomarqueurs d'imagerie peuvent être utilisés pour assurer une prise en charge optimale du patient dans le cadre, par exemple d'une réunion de concertation pluridisciplinaire (RCP). Une réunion RCP est une réunion durant laquelle au minimum trois médecins de spécialités différentes (oncologue, radio-thérapeute et chirurgien) se rencontrent pour discuter sur l'état du patient et proposer une prise en charge thérapeutique qui prend en considération les facteurs de risque (age, sexe, syndromes de prédisposition, maladies auto-immunes, etc.). Le but de la RCP est d'améliorer l'organisation des soins en cancérologie en utilisant entre autres les modèles décisionnels.

#### B. Analyse et "reporting" des images cliniques

La formalisation des structures de stockage des données relatives aux biomarqueurs d'imagerie est indispensable pour faciliter leur bonne réutilisation dans d'autres contextes d'études. La Figure 2.18 décrit le contenu et la structure du compte rendu radiologique des patients de la SEP. Telle que la figure le montre plusieurs biomarqueurs en IRM conventionnelle sont mis en jeu : l'évaluation de la forme des lésions et de leurs tailles, le calcul du nombre de lésions actives, la quantification semi-qualitative de l'atrophie cérébrale, etc.

Sur le marché industriel, on peut citer quelques outils qui automatisent la génération des comptes rendus radiologiques :

- L'outil "Quantib Brain"<sup>10</sup> développé par la société Quantib qui est spécialisée dans la standardisation de l'analyse quantitative des images médicales de types CT et IRM. Cet outil permet de détecter et de mesurer l'évolution longitudinale des valeurs de biomarqueurs d'imagerie pour les pathologies abdominales, cardiovasculaires et neurologiques. La Figure 2.19 illustre un exemple de rapport généré par l'outil pour l'évaluation des paramètres d'imagerie IRM de la sclérose en plaques entre deux temps tout en spécifiant les types de ressources employés dans le calcul.
- L'outil MSmetrix<sup>11</sup> d'icometrix qui aide les radiologues et neuro-radiologues dans l'analyse des images IRM des patients atteints de la sclérose en plaques afin de mieux suivre leur état. Cet outil est basé sur la mesure de l'atrophie du cerveau du patient (perte de volume du cerveau) et la charge de lésions cérébrales. Les résultats de ces analyses quantitatives sont structurées d'une manière synthétique (comme le montre la Figure 2.20) dans un rapport de synthèse qui facilite l'interprétation. Cet outil est approuvé pour la pratique clinique en Europe, au Canada et dans de nombreux autres pays dans le

---

10. <http://www.quantib.com/>

11. <https://icometrix.com/>

**Box 4 | The radiological report**

To facilitate communication between the radiologist and the referring neurologist, radiological reports for patients with suspected or clinically definite MS should contain the following information.

**MRI technique**  
A brief, concise description of the MRI technique should include the anatomical area covered (brain, spinal cord, optic nerve), as well as the field strength, slice thickness, type and dose of contrast agent used, and type of sequences performed. These data are needed for proper comparative analysis of examinations that are performed at different time points and in different imaging centres.

**Findings**  
This section should start with a comprehensive, systematic description of all imaging findings related to the specific clinical situation, using standardized terminology. Examples of such findings include:

- Lesion number (T2 and gadolinium-enhancing T1), topography, size, and shape (with reference to MS characteristics)
- Qualitative assessment of T2 and T1 lesion load
- Semiquantitative visual assessment of brain atrophy
- Positive and negative imaging features that could be considered as evidence for or against the diagnosis of MS
- In follow-up scans, the number of unique active lesions defined as gadolinium enhancing, plus unenhanced new and substantially enlarged T2-hyperintense lesions
- Any incidental or unexpected findings, which should be clearly described and interpreted as either clinically relevant or irrelevant

**Conclusion**  
The report should always contain a conclusion to briefly communicate the radiological interpretation, particularly as related to the clinical problem. Examples of pertinent data include identification of typical or atypical MS lesions, or MS-unrelated lesions and the differential diagnosis thereof; fulfilment of MRI diagnostic criteria for dissemination in space and time; and evidence of disease activity and progression.

Abbreviation: MS, multiple sclerosis.

FIGURE 2.18 – Rapport radiologique pour regrouper les données d'observation issues du diagnostic des patients de la SEP, extraite de [Filippi 2016]

monde entier.

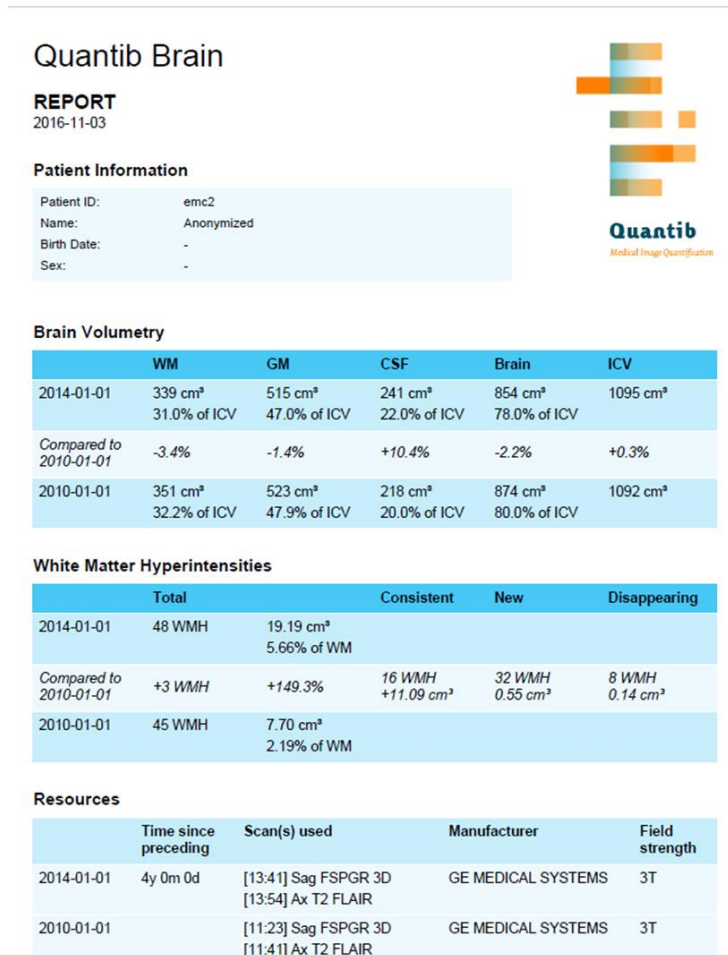


FIGURE 2.19 – Exemple de rapport généré par le "Quantib Brain", extraite de <http://www.quantib.com/>. Le rapport contient des mesures de (1) volume de la matière grise, matière blanche, le liquide cérébro-spinal et du cerveau, (2) de l'hyperintensité dans la matière blanche. L'objet de ce rapport est l'étude de l'évolution de ces mesures entre les deux années 2010 et 2014 pour le patient en question.

## 2.6.2 Description des besoins médicaux dans les applications dédiées à la recherche

Dans cette partie, nous définissons quelques besoins fonctionnels des sites de recherche qui réalisent des études qui sont basées sur les biomarqueurs d'imagerie. Ces études peuvent se répartir en deux catégories : des recherches cliniques et des recherches translationnelles.

### **2.6.2.1 En recherche clinique**

La recherche clinique comprend notamment les essais cliniques effectués chez l'Homme pour soit tester de nouvelles stratégies de soins, soit comparer des stratégies de soins existantes ou des techniques de diagnostic existantes afin de les améliorer. Pour réaliser et faire avancer ces recherches cliniques, la mise à disposition des bases de données cliniques est nécessaire. En cancérologie par exemple, une grande base de données des lésions cibles ( disponible sur le site du "National Cancer Institute" <sup>12</sup>) a servi pour tester l'impact de la modification des critères RECIST 1.1 sur la réponse des patients aux traitements.

En recherche clinique, les sites de recherche ont, essentiellement besoin de :

- Développer et de valider des biomarqueurs d'imagerie pour un rôle clinique précis et une population de patients bien déterminée.
- Valider des algorithmes d'imagerie en réutilisant des données de recherche pour la validation et la qualification des algorithmes d'imagerie quantitative. Les algorithmes de validation des biomarqueurs d'imagerie vont permettre de tester les méthodes de traitement d'images sur différentes modalités d'images, appareils de mesure et entres différents centres de recherche. Les algorithmes de qualification de biomarqueurs d'imagerie vont permettre la validation sur une cohorte de données de patients, des différents modalités d'images et multiples interventions thérapeutiques.
- etc.

### **2.6.2.2 En recherche translationnelle**

La recherche translationnelle sert à favoriser l'exploration de nouvelles pistes, théories ou concepts à partir d'une observation clinique afin de valoriser des découvertes scientifiques. En recherche translationnelle, les centres de recherche réalisent à titre d'exemple des études autour de :

- la compréhension de la signification biologique de biomarqueurs d'imagerie pour la population de patients déterminée par exemple (étude radiogénomique visant à corréliser les valeurs de biomarqueurs d'imagerie à des signatures d'expression de gènes dont on connaît la fonction).
- La conception de nouveaux biomarqueurs d'imagerie à partir de biomarqueurs existants, inclut différents types d'applications telles que : l'application de biomarqueurs existants pour des rôles cliniques similaires mais appliqués à d'autres populations de patients, l'application de biomarqueurs existants pour des rôles cliniques différents mais appliqués aux mêmes populations de patients, l'application de biomarqueurs d'imagerie existants pour des rôles cliniques différents et appliqués à d'autres populations de patients.
- etc.

En résumé, pour supporter les études de recherche utilisant les biomarqueurs d'imagerie les infrastructures informatiques doivent être capables de :

---

12. <https://www.cancer.gov/>

- Calculer des biomarqueurs d'imagerie dans le cadre d'un projet de recherche selon les protocoles définis dans la procédure,
- Sauvegarder des biomarqueurs d'imagerie dans des bio-banques d'images,
- Aider à la découverte des nouveaux biomarqueurs d'imagerie,
- Rechercher des données d'imagerie pour alimenter des projets de recherche,
- Tester l'efficacité de biomarqueurs d'imagerie sur différents bio-banques,
- etc.



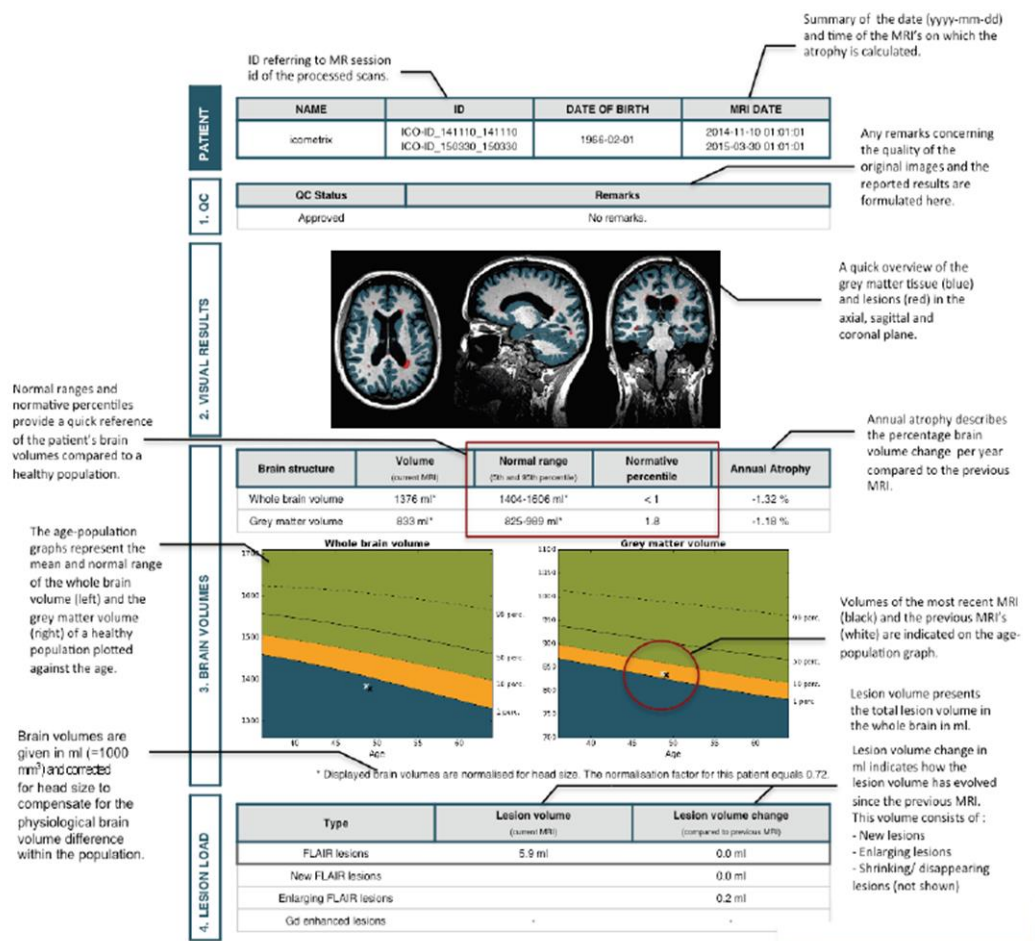


FIGURE 2.20 – Exemple de rapport de MSmetrix, Dans la section "visual results" les points rouge représente les zones d'hypo-intensité et la couleur bleue représente la matière grise. Dans la section "brain volumes" "annual volume change" est le pourcentage de changement du volume du cerveau/ matière grise entre l'année 2015 et 2014. la valeur "normal volume change" représente le changement de volume annuel moyen pour une personne en bonne santé.



## Chapitre 3

# État de l'art de la thèse

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>59</b>
<b>3.2</b>	<b>Représentation de la sémantique des données d'imagerie avec des standards informatiques</b>	<b>60</b>
3.2.1	Comptes rendus structurés avec le standard DICOM	60
3.2.2	Sémantique reposant sur le modèle d'information AIM	67
<b>3.3</b>	<b>Structures de données dans les systèmes à base de connaissances</b>	<b>70</b>
3.3.1	Terminologies, thésaurus et ontologies	70
3.3.2	Quelques standards fondamentaux du Web sémantique	72
<b>3.4</b>	<b>Représentation des connaissances biomédicales</b>	<b>74</b>
3.4.1	Format UMLS	74
3.4.2	Librairie OBO	75
<b>3.5</b>	<b>Représentation de la sémantique des données d'imagerie avec des ontologies</b>	<b>76</b>
3.5.1	Le modèle Open Annotation	76
3.5.2	L'ontologie OntoNeuroLOG	78
3.5.3	L'ontologie QIBO	79
3.5.4	Le Système BiomRKRS	81
3.5.5	Une représentation des biomarqueurs fondée sur l'ontologie OGMS	84
3.5.6	Le projet THESEUS MEDICO	84
<b>3.6</b>	<b>Discussion</b>	<b>87</b>
3.6.1	Problématique générale des contributions de la thèse	87
3.6.2	Contributions de la thèse	88

---

### 3.1 Introduction

Dans le chapitre précédent, nous avons décrit le contexte général de la thèse et nous avons défini les principaux besoins liés au partage des données d'imagerie

et introduit la problématique des biomarqueurs d'imagerie. Dans ce chapitre, nous présentons les principaux formats informatiques (modèles d'information, terminologies, ontologies, etc.) existants pour la représentation de la sémantique des données d'imagerie.

Ce chapitre est organisé comme suit : la Section 3.2 présente le standard DICOM et le modèle AIM (Annotation and Image Markup). La Section 3.3 décrit les différentes structures de données disponibles pour l'enrichissement sémantique des données. La Section 3.4 donne des exemples sur les formats de données les plus adoptés dans l'implémentation des systèmes à base de connaissances dans le domaine biomédical. Dans la Section 3.5, nous présentons les solutions existantes pour la description sémantique des données d'imagerie. Enfin, nous clôturons ce chapitre par une section Discussion dans laquelle nous énonçons la problématique de la thèse et nous situons brièvement nos contributions.

## 3.2 Représentation de la sémantique des données d'imagerie avec des standards informatiques

### 3.2.1 Comptes rendus structurés avec le standard DICOM

Le *standard DICOM* (Digital Imaging and Communications in Medicine) a défini une structure de données appelée *compte rendu structuré DICOM* (en anglais DICOM structured reporting ou DICOM SR) [Clunie 2000] couvrant le domaine des observations d'imagerie incluses dans les rapports cliniques. Les observations d'imagerie concernent principalement les modalités d'images DICOM, les images dérivées, les résultats de segmentation, les mesures (taille, volume, surface, etc.), les cartes paramétriques, etc. Cette standardisation de la description des observations d'imagerie repose sur la définition des "SR templates", que l'on peut littéralement traduire en français en "modèles SR". Les modèles SR contraignent l'organisation des données en spécifiant les concepts et les relations codés ainsi que leurs significations.

Pour afficher d'une manière intelligible les comptes rendus SR un toolkit spécifique (par exemple basé sur le DICOM toolkit DCMTK<sup>1</sup> doit être implémenté pour les stations de PACS (Picture Archiving and Communication System) afin de présenter les objets de l'arbre SR, par exemple dans le format HTML (HyperText Markup Language) [Hussein 2004]. Les données de DICOM SR sont sérialisées en syntaxe DICOM (triplet, <tag, longueur, valeur>) et organisées sous forme d'arbre. La hiérarchie de la structure arborescente ainsi que les spécifications des comptes rendus DICOM ont été décrites en détail dans les publications de David Clunie [Clunie 2000] et Noumeir [Noumeir 2003].

#### 3.2.1.1 La terminologie dans DICOM SR

Les nœuds contenus dans DICOM SR sont dans la plupart des cas codés à l'aide de ressources terminologiques externes, SNOMED (Systematized Nomenclature of

1. [http://support.dcmk.org/docs/mod\\_dcmsr.html](http://support.dcmk.org/docs/mod_dcmsr.html)

Medicine) et Radlex (Radiology Lexicon) sont les deux principales terminologies employées par le standard DICOM. Les termes retenus de ces différentes ressources externes sont listés dans la partie 16 du standard<sup>2</sup>. Les spécifications DICOM SR utilisent également des concepts d'imagerie qui proviennent du standard DICOM dans les cas où les termes proposés par les lexiques existants ne sont pas très pertinents pour la modélisation de certaines informations. Les concepts qui sont inclus dans la ressource terminologique propre à DICOM est appelée "DICOM controlled terminology" ou DCT.

**SNOMED** SNOMED [SNOMED 2015] est une terminologie médicale qui représente les principaux concepts de la médecine clinique. Les concepts sont répartis en 19 axes sémantiques. Les concepts sont représentés par une série de termes (synonymes, termes préférés, etc.) au niveau de chaque axe sémantique. SNOMED CT<sup>3</sup> (SNOMED Clinical Terms) est l'une des terminologies cliniques les plus complètes pour toutes les spécialités médicales (comporte environ 350 000 concepts cliniques). Cette terminologie est utilisée pour décrire des informations médicales (l'anatomie, les maladies, les interventions chirurgicales, les diagnostics, les résultats et les décisions thérapeutiques, etc.) dans les dossiers électroniques des patients (voir Figure 3.1). Les concepts sont inter-liés via des relations; une relation de type "est un" pour spécifier un concept générique et une deuxième relation dite d'attributs pour lier les concepts entre eux (comme par exemple, "SNOMEDCT:finding site", "SNOMEDCT:causative agent", etc.). Elle est disponible en plusieurs langues et elle est utilisée sous l'obtention d'une licence dans plus de 50 pays. La principale limite de SNOMED CT est qu'elle n'est pas librement accessible, une licence doit être achetée pour avoir le droit d'utilisation de ce lexique.

**Radlex** Radlex<sup>4</sup> est une terminologie qui couvre le domaine de la radiologie, développée par le RSNA (Radiological Society of North America). La terminologie Radlex décrit des termes radiologiques variés : observations d'imagerie, résultats cliniques, modalités d'imagerie, entités anatomiques etc. La Figure 3.2 montre la hiérarchie globale de Radlex<sup>5</sup>. Radlex inclut des concepts existants dans SNOMED [SNOMED 2015] et DICOM. Son objectif est d'améliorer la recherche et la communication des données cliniques d'imagerie.

Cette terminologie est largement employée par les radiologues. Elle a été utilisée dans des applications cliniques pour représenter des résultats issus des procédures radiologiques [Kahn Jr 2009] et pour indexer des comptes rendus radiologiques en se basant sur le contenu [Lacson 2012]. Certains problèmes ont été détectés dans Radlex, dont beaucoup pourraient être rectifiés : la plupart des termes sont entièrement sans définitions et d'autres sont mal classés (par exemple, le terme "Radlex:imaging observation" est formellement défini comme une sous classe de "Radlex:Radlex

---

2. <http://dicom.nema.org/medical/dicom/current/output/html/part16.html>

3. <http://www.ihtsdo.org/snomed-ct/>

4. <http://www.radlex.org>

5. <http://playbook.radlex.org>

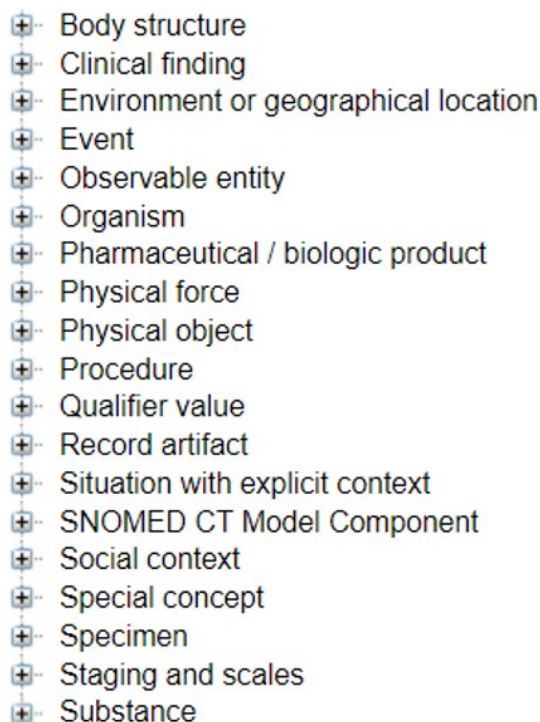


FIGURE 3.1 – Structure globale de la terminologie SNOMED CT

`entity`" et textuellement définie comme une `"Radlex:property"` ("Features seen and reported by a radiologist during interpretation of an image that are specific to imaging"), ce qui est sémantiquement incorrect. Le terme `Radlex:ADC measurement` est défini comme un type de propriété qui est de type `"Radlex:measurement type"` et non pas comme une valeur de mesure d'une propriété.). En 2007, une « ontologisation » de cette ressource a été proposée pour sémantiquement l'enrichir [Rubin 2008a].

De notre point de vue, nous pensons que la terminologie Radlex doit être améliorée pour faciliter son exploitation par les utilisateurs. Les définitions des termes sont décrites sous la forme de propriétés et non d'annotations comme on le fait le plus souvent, ce qui fait qu'elles ne sont visibles qu'au niveau des instances. Cela tient au fait que RadLex est représentée en OWL FULL où les entités sont à la fois classes et instances, ce qui ne simplifie pas pour l'utilisateur la lecture des informations comme les définitions ou les 'preferred labels'.

### 3.2.1.2 La syntaxe et les règles dans DICOM SR

Le standard DICOM définit différents types de noeuds SR qui sont listés ci-dessous :

- Noeud de type `"CONTAINER"`, utilisé pour structurer les données dans le document SR ;

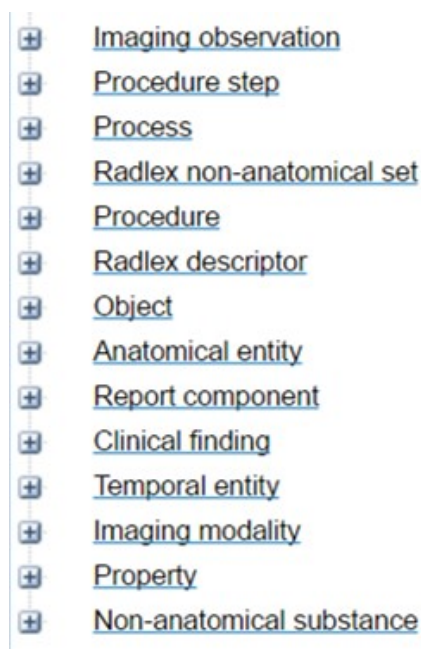


FIGURE 3.2 – Structure globale de la terminologie Radlex

- Nœud de type "NUM", qui représente les mesures composées de valeurs numériques accompagnées de leurs unités ;
- Nœud de type "TEXT", qui décrit des observations radiologiques sous forme de texte libre ;
- Nœud de type "CODE", qui est utilisé pour représenter les concepts codés et leurs valeurs définies dans la partie 16 de la norme DICOM. Par exemple, le concept de DCM "Single Image Finding" (valeur du code 111059) peut avoir comme valeur du code du concept de SRT "Mammography breast density" F-01796, ;
- Nœud de type "SCoord", (spatial), qui décrit des formes graphiques qui peuvent être des points, un ensemble de points ou d'autres formes graphiques de type cercle, ellipse ou poly-line, etc. Ce type de nœud est utilisé pour définir une région d'intérêt sur une séquence d'image DICOM. Il existe aussi un 3D SCoord qui est utilisé lorsque les graphiques sont en forme 3D ;
- Nœud de type IMAGE fait référence via l'utilisation d'un identifiant unique à l'image médicale étudiée ;
- Nœud de type "TCoord" (temporel), qui se réfère aux régions d'intérêt temporelles qui sont suivies sur une série d'images DICOM ;
- Nœud de type WAVEFORM", qui fait référence à un signal qui représente certaines qualités physiques (par exemple, tension électrique, pression, gaz concentration ou son). La qualité mesurée peut concerner l'anatomie du patient, la thérapeutique dispositifs, etc. ;
- Nœud de type "PNAME" (nom de personne), utilisé pour définir le nom de

- l'observateur ou du patient ;
- Nœud de type "DATE", "HEURE" utilisés respectivement pour représenter la date et l'heure de création/ modification/ suppression d'un objet d'information ;
- Nœud de type "DATETIME", qui combine les valeurs des attributs DATE et TIME ;
- Nœud de type "UIDRE", qui représente des identifiants uniques associés à des entités DICOM ;

Les relations au sein d'un compte rendu SR servent à établir des liens entre les différents nœuds : la relation "HAS PROPERTIES" décrit les propriétés du nœud parents ; la relation "INFERRED FROM" spécifie ce qui a conduit à l'inférence considérée ; la relation "CONTAINS" associe les nœuds de type CONTAINER avec leurs nœuds descendants et la relation "SELECTED FROM" lie le nœud de type SCOOD avec le nœud fils, à savoir un nœud image dont la région d'intérêt est extraite.

### 3.2.1.3 Codage des résultats d'imagerie avec DICOM SR

Les modèles SR définissent la structure d'un document SR pour un usage particulier ; les spécifications portent sur les nœuds et les relations SR, les codes des concepts et l'ensemble de valeurs qu'ils peuvent recevoir, etc. Deux types de modèles SR existent :

- le modèle de racine ("root template SR" en anglais) à la représentation d'un document complet (par exemple, "chest CAD") ;
- le sous-modèle ("sub-template" en anglais) qui porte sur un sous-ensemble, il est en général générique et réutilisable dans plusieurs types de comptes rendus (par exemple, "linear measurement template").

La structure CAD (*Computer Aided Detection*) SR stocke les critères quantitatifs et qualitatifs d'imagerie, la Figure 3.3 illustre la structure générale d'un compte rendu d'analyse CAD d'une mammographie. Elle montre une observation radiologique de type "mammography breast density" associée à une région d'intérêt.

Les modèles SR des CAD, et plus généralement les DICOM SR ont fourni une base intéressante pour structurer les informations qualitatives et quantitatives liées aux images DICOM. Mais ces derniers doivent être étendus pour répondre aux besoins spécifiques des applications de la recherche clinique. David Clunie et al. ont proposé et développé dans le projet QIICR (Quantitative Image Informatics for Cancer Research) un format de stockage (codes DICOM et modèles) des résultats des essais cliniques en oncologie avec le standard DICOM SR [Clunie 2007b]. Les principaux concepts identifiés et codés en DICOM SR afin de bien décrire les essais cliniques sont :

- Observateur : le contexte de l'observateur (utilisateur, rôle, ID, etc.) est inclus au niveau supérieur du document SR ;
- Sujet : patient, animal ou spécimen ;
- "endpoint" : un ou plusieurs "endpoint" dans une instance de document



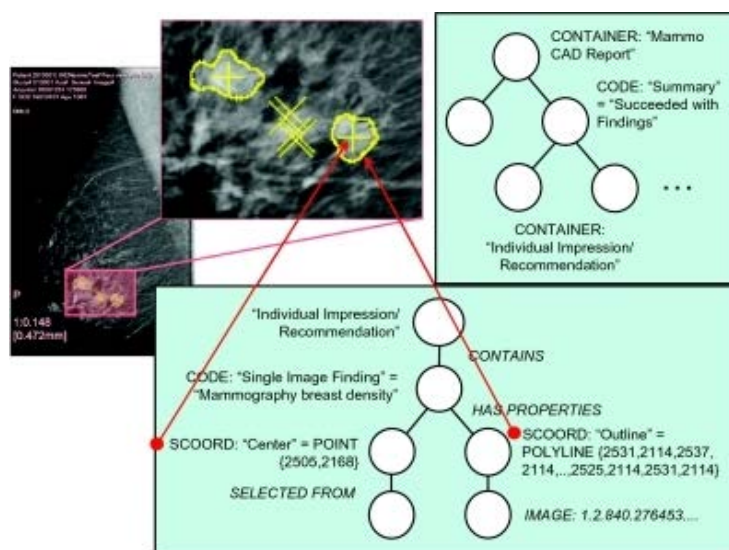


FIGURE 3.3 – CAD d'une mammographie, les résultats d'imagerie sont codés sous le format DICOM SR [Clunie 2007a]

SR sont décrits séparément à partir des sous-tâches (par exemple, plusieurs "endpoints" peuvent être modifiés dans une seule sous-tâche) ;

- Étude : se réfère à des études d'imagerie à partir de différents moments utilisés lors d'une sous-tâche d'une seule lecture ;
- Lésion : utilisé pour délimiter des régions d'intérêt ;
- Objet de mesure : utilisé pour définir des lésions et des emplacements pour des mesures supplémentaires de lésions ;
- Observation : le terme « observation » comprend les mesures, les évaluations de réponse thérapeutique ainsi que les commentaires créés pour un "endpoint", par un seul observateur et pour un sujet unique ;
- "Clinical Trial Audit Trail" : permet de gérer les différentes versions de documents SR (pointeur de référence vers les versions précédentes de la version actuelle d'un document DICOM SR). Il stocke aussi les informations de la création : (qui ? quand ? quoi ?) et de l'édition (qui ?, quoi ?, quand ?, pour-quoi ?) de chaque changement de données.

Les sous-tâches mises en jeu dans le cadre d'un essai clinique peuvent être de type : localisation (organe, lésion, tissu, etc.), catégorisation (lésion cible, lésion non cible, etc.), description (irrégularité, prise de contraste, etc.), mesure (taille, intensité, texture, etc.), adjudication : comparaison et arbitrage entre deux résultats cliniques réalisés par différents observateurs. Ce travail propose d'agrèger les résultats des essais cliniques des différents "time points" dans une seule instance de rapport SR. L'objectif de ce travail est d'assurer une vue "unique" des instances d'observations d'imagerie.

### 3.2.1.4 Utilisation de DICOM SR en pratique clinique et en recherche

Les avantages de l'utilisation des comptes rendus structurés DICOM SR en pratique clinique sont les suivants :

- Chaque mesure a une unité et une sémantique précise ; la mesure peut être décrite explicitement, en utilisant une relation de référence directe à l'image ou en indiquant simplement la signification de la mesure (diamètre, volume, longueur, etc.) ;
- la recherche basique de l'information peut être faite à travers le parcours des nœuds "résultats" ;
- les images peuvent être référencées dans le document en incluant leur UID dans un nœud. De plus, on peut associer des régions spatiales à une ou plusieurs images pour désigner des sites anatomiques ou des lésions ;
- la sécurité et la cohérence des informations relatives à l'étude sont assurées ;
- Les comptes rendus collaboratifs impliquant plusieurs observateurs (humains/ dispositifs) sont traités ;
- DICOM SR spécifie la relation entre les observations faites, les images associées et le diagnostic retenu ;
- etc.

Malgré la large utilisation du standard DICOM en pratique clinique, la structure DICOM SR reste globalement peu utilisée pour des raisons expliquées dans [Bosmans 2012]. DICOM SR est une structure adaptée à l'échange, mais inapte au traitement de requêtes ou inférences ; ceci est dû à l'absence d'un cadre formel pour représenter les contraintes liées aux différents types de nœuds et de relations ainsi que la non exploitation de la signification ontologique des codes utilisés (e.g. SNOMED). Jusqu'à aujourd'hui, il n'y a pas d'ontologie DICOM, même si certains travaux de recherche se sont intéressés à la question de la représentation sémantique du standard DICOM. On peut citer à titre d'exemple l'ontologie "Semantic Digital Imaging and Communications in Medicine Ontology" <sup>6</sup> disponible sur le Bioportal ou bien la transformation en format RDF des méta-données DICOM qui est proposée dans [Brunnbauer 2013].

Le travail de Mhiri et al. [Després 2007] propose un outil d'annotation des comptes rendus DICOM SR afin de faciliter leur indexation et recherche. L'outil est basé sur une approche ontologique de modularisation du contenu informationnel selon six axes sémantiques qui concernent le contexte, les descripteurs visuels, les techniques, l'anatomie, les pathologies et les recommandations. La Figure 3.4 illustre ces différents axes de la description des résultats d'observation d'imagerie. De notre point de vue, l'une des principales difficultés de la reconstruction des comptes rendus DICOM SR autour d'une ontologie réside dans le référencement d'autres ressources terminologiques en particulier SNOMED et RadLex. Les ontologies Radlex et SNOMED CT proposées sont de grande taille (SNOMED dépassent les 300 000 concepts et Radlex dépassent les 75 000) et elle doivent être révisées pour être alignées sur une ontologie fondatrice et réutiliser par d'autres ontologies de domaines existants.

6. <https://wiki.nci.nih.gov/display/Imaging/DICOM+Ontology>

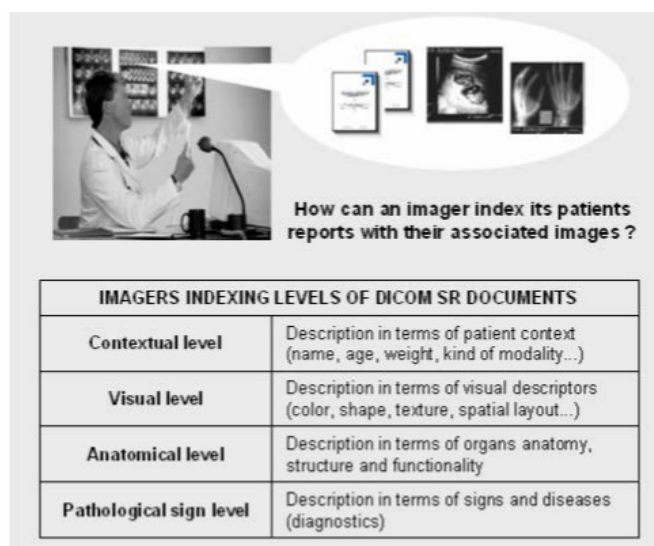


FIGURE 3.4 – Les axes sémantiques de l’annotation des contenus informationnels des DICOM SRs, extraite de ce papier [Després 2007]

### 3.2.2 Sémantique reposant sur le modèle d’information AIM

Le modèle d’information AIM (*Annotation and Image Markup*) a été développé dans le cadre de l’initiative caBIG (Cancer Biomedical Informatics Grid) imaging [Channin 2010]. Ce modèle est basé sur la représentation des annotations radiologiques qui se réfèrent à des mesures, des textes, des observations, des formes graphiques délimitant des régions d’intérêt, etc (voir Figure 3.6). Ces annotations sont catégorisées en cinq groupes : le groupe de calcul, le groupe de contenu sémantique de l’image, le groupe de «Mark up», le groupe de références vers les images et le groupe des informations générales. Ici, on ne décrit que les trois groupes suivant :

- **Groupe calcul** : ce groupe représente les résultats de calcul décrits dans des annotations AIM qui peuvent être associées ou pas à des formes graphiques. C’est au niveau de ce groupe qu’on spécifie la méthode de calcul (segmentation) qui a été exécutée en spécifiant les paramètres (positions des pixels) ainsi que les algorithmes mis en jeu.
- **Groupe "mark up"** : ce groupe modélise des informations textuelles ou des représentations graphiques (référéncées avec SCOORD dans les DICOM SRs) faites par des observateurs. Les formes graphiques qui sont introduites dans le modèle sont de différents types : point, multipoint, polyline, cercle et ellipse ; des coordonnées dans les espaces 2D et 3D leurs sont associées. Ici, on s’intéresse aux formes qui permettent la délimitation de la région d’intérêt mais pas à la nature de la qualité mesurée (la qualité mesurée peut être 2D (par exemple, une surface) malgré le fait que la forme utilisée soit 3D).
- **Groupe sémantique de l’image** : ce groupe permet de représenter les observations cliniques tout en faisant la distinction entre le monde réel (l’ob-

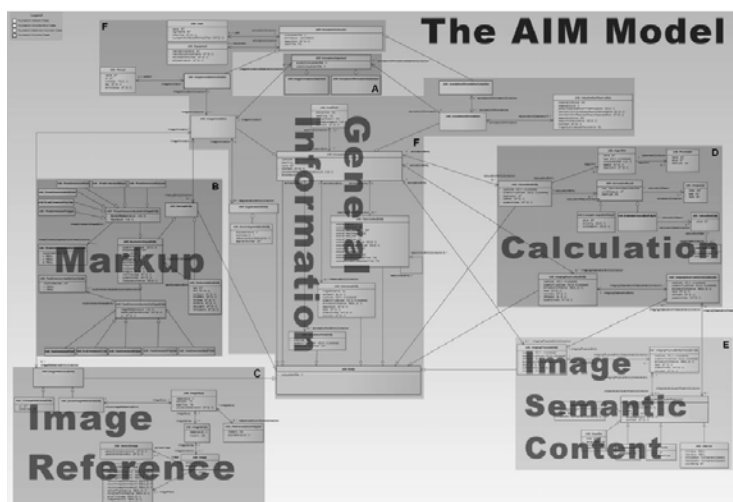


FIGURE 3.5 – Les groupes d’annotation du modèle AIM [Channin 2010]

jet physique : e.g., la masse chez le patient) et le monde de l’information (contenu informationnel : la masse observée sur l’image médicale). Ainsi le groupe distingue deux classes élémentaires liées au contenu de l’image médicale : "AIM : physical imaging entity" qui désigne un référent (par exemple la masse présente dans le corps du patient) et "AIM : image viewing entity" qui représente des références (exemple de la masse observée sur l’image médicale). On a dans le modèle deux types de caractéristiques celles liées à l’objet (par exemple, dilaté ou pas) et celles qui font référence à l’apparence de l’objet physique vu sur l’image, qui peut être réelle ou non (artéfact).

Les données du modèle AIM peuvent être stockées dans un fichier XML. Des outils open source existent pour sérialiser les données sous le format AIM, on peut citer l’API (Application Programming Interface) AIM<sup>7</sup> qui facilite l’intégration d’AIM dans des applications d’annotation, l’outil ePad (Electronic Physician Annotation Device) [Rubin 2008b], les référentiels de méta-données d’image implémentant AIM comme par exemple le Biomedical Image Metadata Manager BIMM [Korenblum 2011] ; Le modèle AIM a été utilisé pour le suivi longitudinal des mesures de lésions dans les séries d’imagerie dans les essais cliniques, par exemple, dans ePAD [Rubin 2014], les lésions ont été enregistrées comme annotations d’image AIM et étiquetées par un identifiant unique.

Le modèle AIM a introduit les principales entités mises en jeu dans l’annotation d’image, mais sa mise en œuvre manque de sémantique formelle, car elle n’est pas basée sur des ontologies. De ce fait, nous ne pouvons pas représenter des entités complexes ou effectuer un raisonnement basé sur la logique pour inférer de nouvelles connaissances sur le contenu de l’image.

En résumé, l’annotation sémantique des données d’imagerie concerne à la fois les données d’entrées et de sorties résultant de l’exécution du processus d’acquisition,

7. [https://web.stanford.edu/group/qil/cgi-bin/mediawiki/index.php/AIM\\_API/](https://web.stanford.edu/group/qil/cgi-bin/mediawiki/index.php/AIM_API/)

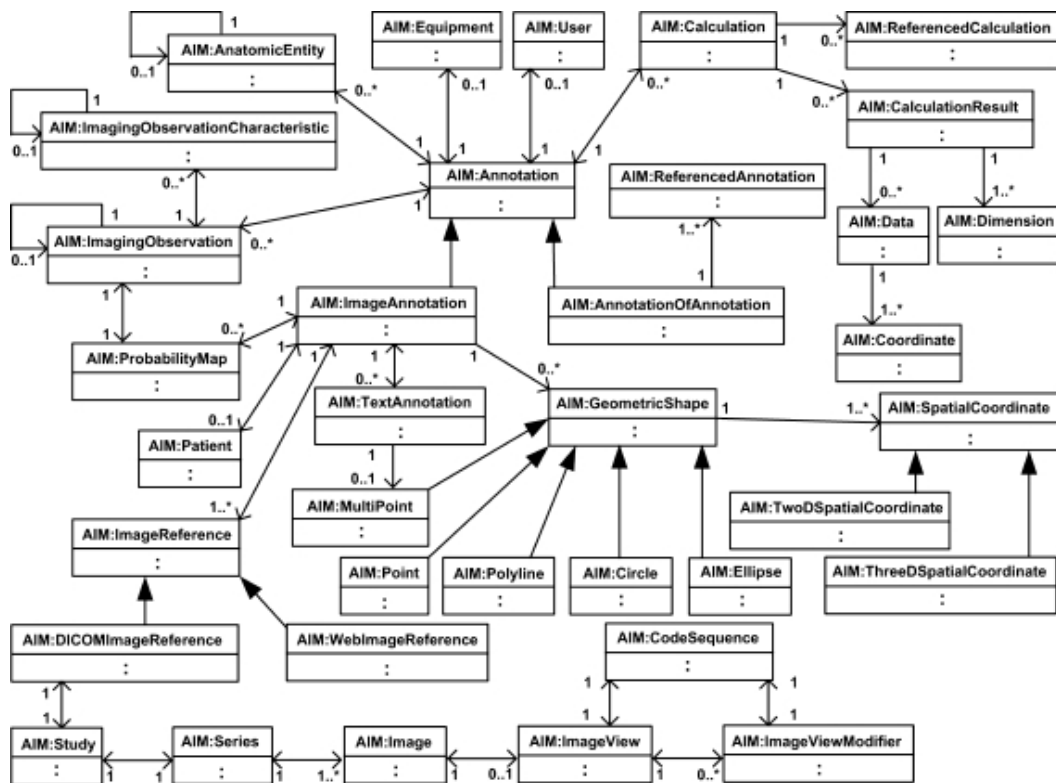


FIGURE 3.6 – Le diagramme UML du modèle AIM [Channin 2010]

d'analyse des données, de visualisation et de création des rapports. L'attribution d'une sémantique précise aux phénotypes d'images (ensemble de traits observables d'un individu), va permettre de partager d'une façon fiable les informations sur leur rôles cliniques, leur degré de validation ainsi que sur les contextes cliniques dans lesquels ils ont été validés. Tous ceci contribue directement à une utilisation sûre des biomarqueurs d'imagerie comme des outils d'aide à la décision ainsi que la création d'un réseau de connaissances sur une certaine pathologie pour stocker les données moléculaires, biologiques, etc. Plusieurs travaux [Ruttenberg 2007] [ESR 2015] ont souligné que la structuration des données d'imagerie avec des simples terminologies ou modèles d'information n'est pas suffisante et que l'enrichissement sémantique de ces données peut apporter des fonctionnalités complémentaires. Dans le reste de ce chapitre, nous nous focalisons sur les systèmes qui annotent sémantiquement les données d'imagerie.

### 3.3 Structures de données dans les systèmes à base de connaissances

Dans cette section, nous présentons les différentes structures de données qui sont adoptées dans l'implémentation des systèmes à base de connaissances du domaine biomédical. Ces systèmes ont comme principal objectif de supporter les activités des scientifiques et des cliniciens.

#### 3.3.1 Terminologies, thésaurus et ontologies

Actuellement, trois structures de données sont utilisées dans les systèmes à base de connaissances liés au domaine biomédical : les terminologies, les thésaurus et les ontologies. Cette sous-section est dédiée à la définition de ces différentes structures, qui donnent lieu à des confusions fréquentes [Schulz 2009].

##### 3.3.1.1 Terminologie

Une terminologie est un vocabulaire qui est composé d'un ensemble fixe de termes utilisés dans un domaine spécifique. Les terminologies attribuent des définitions textuelles pour décrire la sémantique des termes qu'elles contiennent. Ainsi, la sémantique n'est compréhensible que par les humains. Les terminologies sont utilisées pour établir une compréhension commune d'un ensemble de termes. On parle d'une taxonomie (ou bien une hiérarchie de type) quand la terminologie est structurée par hiérarchie. La hiérarchie des termes peut être représentée à l'aide de la relation "SKOS :broader than" qui traduit que le terme ainsi tous ses descendants sont sous une autre catégorie.

### 3.3. Structures de données dans les systèmes à base de connaissances71

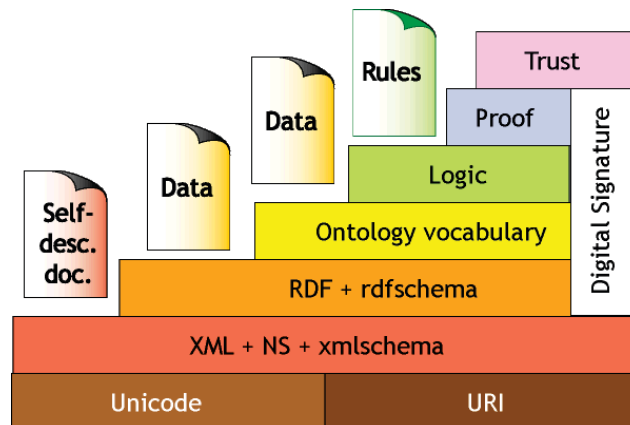


FIGURE 3.7 – L'infrastructure des technologies du Web sémantique, d'après T. Berners Lee

#### 3.3.1.2 Thésaurus

Un thésaurus est une terminologie où les termes sont en plus regroupés par leur similitude sémantique. L'objectif d'un thésaurus est de constituer un vocabulaire normalisé et d'organiser la liste des termes de ce vocabulaire sans forcément les définir, dans le but notamment d'indexer un corpus de ressources documentaires et de faciliter les recherches dans ce corpus. Le thésaurus permet à la fois de guider l'indexation en suggérant des termes à utiliser et d'enrichir des requêtes ou de proposer des reformulations. L'exemple d'un thésaurus biomédical est le thésaurus du NCI (National Cancer Institute thesaurus).

#### 3.3.1.3 Ontologie

Le terme "ontologie" est originaire de la philosophie et désigne «la philosophie de l'être». Il a été utilisé en informatique et plus spécifiquement dans le domaine de l'intelligence artificielle pour décrire une organisation hiérarchique basée sur des concepts et des relations qui sont exprimés dans un langage formel. L'ontologie est définie par Thomas Gruber en 1993 comme "une spécification formelle et explicite d'une conceptualisation partagée" [Gruber 1993]. Contrairement aux terminologies la sémantique des classes d'une ontologie peut être appréhendée par les machines. De plus, au-delà de la relation de subsumption, d'autres propriétés existent dans les ontologies. Par exemple, la relation "partie-de" (tout organe anatomique humain est une partie d'un corps humain) est une connaissance ontologique.

La différence principale entre un thésaurus et une ontologie est qu'un thésaurus est un recueil de termes décrits en langue naturelle et qui organise des termes et non des catégories, des concepts ou des classes comme c'est le cas dans une ontologie.

Guarino propose de catégoriser les ontologies selon le niveau de dépendance à un point de vue particulier [Guarino 1998]. Il distingue principalement : les ontologies de haut niveau, les ontologies de domaine et les ontologies d'application.

- Les ontologies de haut niveau appelées aussi ontologies fondamentales fournissent les concepts de base pour la représentation du savoir indépendamment d'un domaine particulier. Les entités de base d'une ontologie de niveau supérieur sont les objets, les qualités, les processus, etc.
- Les ontologies de domaine décrivent le vocabulaire associé à un domaine de portée générique (tel que la médecine par exemple on n'y trouve des concepts de diagnostic, structure anatomique et des relations comme celles liées à la localisation d'une structure pathologique dans un organe); ceci se fait en spécialisant les termes introduits dans l'ontologie de haut niveau.
- Les ontologies d'application spécifient un modèle particulier d'un domaine selon le point de vue des professionnels.

### 3.3.2 Quelques standards fondamentaux du Web sémantique

Nous décrivons ici quatre recommandations du Web sémantique (W3C<sup>8</sup>) qui sont : RDF (Resource Description Framework), RDFS (RDF schema), OWL (Ontology Web Language) pour la représentation des données, et SPARQL (SPARQL Protocol and RDF Query Language) pour l'interrogation des données RDF.

La Figure 3.7 montre que les ontologies utilisent le modèle RDF pour la représentation des données, le schéma RDF pour contraindre les relations entre concepts, le langage OWL pour exprimer et représenter des connaissances complexes du modèle, et le langage SPARQL<sup>9</sup> pour récupérer et manipuler des données stockées au format RDF. Les sous-sections suivantes détaillent ces différents langages.

#### 3.3.2.1 Formats de description des données du Web sémantique

Le langage RDF (Resource Description Framework)<sup>10</sup> est un cadre de référencement des ressources du Web et des liens entre elles. Dans ce vocabulaire de données, les ressources sont identifiées par les URIs "Uniform Resource Identifier". Ces ressources sont décrites en RDF sous la forme de triplets (sujet, prédicat, objet). Les triplets RDF peuvent être sérialisés en XML avec la syntaxe RDF/XML.

Le schéma RDF (en anglais RDF schema) est un langage qui permet de décrire des vocabulaires, des propriétés et des classes de ressources dans le modèle RDF. Le schéma RDF est une extension sémantique du langage RDF, c'est une recommandation proposée début 1999. Ce langage permet de spécifier des ontologies dites légères, c'est à dire de nommer des classes et des propriétés, de donner la signature de ces propriétés (les classes des ressources reliées par ces propriétés) et de définir une organisation hiérarchique de ces classes et propriétés (rdfs : subclassOf et rdfs : subPropertyOf). Ces primitives sont des liens de "spécialisation" ou "est un". Elles permettent aux classes et aux relations d'hériter des caractéristiques définies dans les classes (ou les relations) parentes.

---

8. <https://www.w3.org/>

9. <http://www.w3.org/TR/rdf-sparql-query/>

10. [www.w3.org/TR/2002/WD-rdf-concepts-20021108/](http://www.w3.org/TR/2002/WD-rdf-concepts-20021108/)



### 3.3. Structures de données dans les systèmes à base de connaissances 73

OWL est l'abréviation de "Ontology Web Language"<sup>11</sup>, c'est un langage de description des ontologies dites lourdes qui est basé sur RDF/RDFS. Il enrichit le modèle du schéma RDF en définissant un vocabulaire plus expressif. Le langage OWL permet, en plus des primitives RDF(S), des relations entre les classes (disjonction, intersection, union, etc.), la cardinalité, l'égalité, etc. Le langage OWL permet la définition des types de propriétés, des caractéristiques des propriétés (par exemple la symétrie et la transitivité) et des classes énumérées. OWL est une recommandation du W3C depuis 2004 et OWL2 une recommandation du W3C depuis 2009.

Le langage d'ontologies OWL est décliné en trois sous-langages possédant une expressivité ascendante à savoir : OWL-Lite, OWL-DL (Ontology Web Language-Description Logic) et OWL-Full. La raison principale de cette division concerne la complexité, la calculabilité et l'implémentation du langage.

- Le sous-langage OWL-Lite possède la complexité formelle la plus basse et une expressivité minimale. Il est dédié pour la représentation des thésaurus, des taxonomies ou hiérarchie de classes avec des contraintes simples.
- Le sous-langage OWL-DL est une version décidable du langage informatique OWL. Le sous-langage OWL-DL est recommandé pour la représentation d'ontologies nécessitant une puissance d'expressivité tout en garantissant la décidabilité.
- OWL-Full est une version indécidable du langage informatique OWL. Il est considéré comme une extension du langage RDF, tandis que les deux sous-langages OWL-Lite et OWL-DL sont considérés comme des extensions restreintes de RDF.

#### 3.3.2.2 Interrogation des données du Web sémantique

SPARQL (Simple Protocol and RDF Query Language) fournit le langage d'interrogation des graphes RDF. SPARQL est à RDF ce que SQL est aux bases de données relationnelles, il permet de chercher des motifs de graphes ("graph patterns") ainsi que leurs conjonctions et leurs disjonctions. La description complète de la syntaxe peut être trouvée à l'adresse suivante<sup>12</sup>. La structure de la requête SPARQL est très similaire à la celle de SQL (SELECT, WHERE, GROUP BY, etc.) (voir Liste 3.1).

Listing 3.1 – la clause SELECT permet de choisir les variables ?v du résultat, parmi les variables de la clause WHERE. La clause WHERE contient des triplets qui définissent le motif recherché.

```
SELECT ?v1 ?v2 ... ?vn
WHERE
{
  (sujet1 | vi) (predicat1 | vj) (objet1 | vk).
  ...
  (sujetx | va) (predicaty | vb) (objetz | vc).
}
```

11. <http://www.w3.org/TR/owl-features>

12. <http://www.w3.org/TR/rdfsparql-query/>

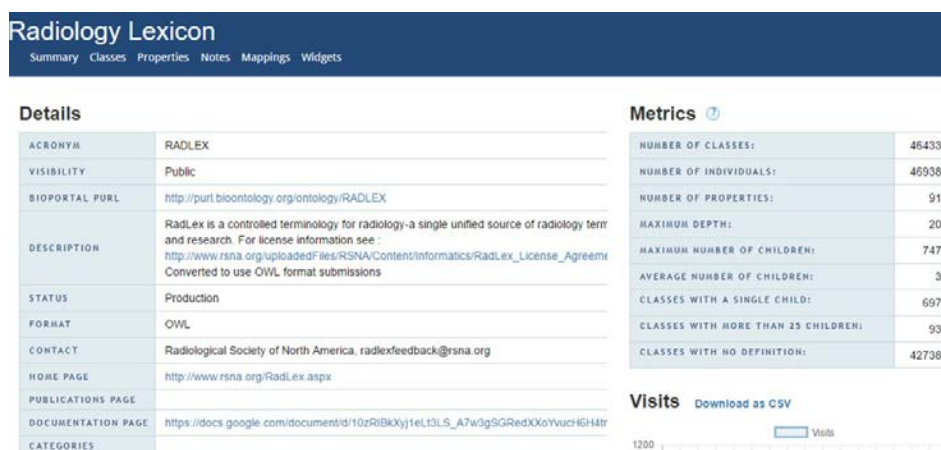


FIGURE 3.8 – Visualisation de l'ontologie Radlex sous le Bioportal NCBO

On note que le langage SPARQL en lui-même ne permet pas de faire des inférences sur les données.

### 3.4 Représentation des connaissances biomédicales

Le NCBO (*National Center for Biomedical Ontology*) est un consortium de biologistes, de cliniciens et d'informaticiens qui vise à supporter la science biomédicale ainsi que les soins cliniques. Cette initiative a fourni aux chercheurs du domaine de la médecine et de la biologie des outils qui leur facilitent l'accès, l'analyse et l'intégration des ressources ontologiques hétérogènes dans le domaine biomédical, à savoir la bibliothèque OBO (Open Biological and Biomedical Ontologies), les dépôts OBD (Open Biomedical Data) et le Bioportal du consortium NCBO [Noy 2009]. La Figure 3.8 montre l'interface Web du Bioportail qui est dédié à la recherche des ressources ontologiques. Cette ressource regroupe principalement des ontologies, des modèles d'information et des terminologies. Le Tableau 3.1 liste quelques exemples de ressources disponibles via ce portail. On note que de nombreuses «ontologies» biomédicales sont plutôt des terminologies, des taxonomies, des thésaurus ou des systèmes de classification et non pas des ontologies.

Dans le domaine biomédical, il existe deux grandes familles de formats de représentation des données : UMLS (Unified Medical Language System) et OBO.

#### 3.4.1 Format UMLS

La bibliothèque UMLS<sup>13</sup> est une ressource de connaissances développée par la NLM aux USA (National library of Medicine) pour répondre aux besoins des applications biomédicales. UMLS a été conçu avant l'émergence des ontologies. Cette bibliothèque combine des terminologies et des systèmes de classification et elle est com-

13. <http://umlsks.nlm.nih.gov>

posée de trois composantes : (1) le méta-thésaurus qui aligne les différents concepts afin de construire un vocabulaire cohérent ; (2) le réseau sémantique qui catégorise les concepts du méta-thésaurus (les principaux groupements des types sémantiques incluent ceux des organismes, les structures anatomiques, les fonctions biologiques, etc.) et spécifie l'ensemble de relations qui les relie (il est considéré comme une ontologie du domaine biomédical). (3) le lexique qui définit syntaxiquement et morphologiquement des concepts qui n'existent pas dans le méta-thésaurus.

Dans [Kumar 2003], les auteurs montrent comment la prise en compte des principes ontologiques permet la construction des ontologies biomédicales plus réalistes et plus robustes vis-à-vis des applications associées.

### 3.4.2 Librairie OBO

La librairie OBO [Smith 2007] a comme objectif de créer des vocabulaires partageables et contrôlés et d'assurer leur évolution coordonnée. La principale différence entre UMLS et la librairie OBO est que si l'objectif de la bibliothèque UMLS est d'intégrer les terminologies existantes et les systèmes de codage, la bibliothèque OBO assure l'évolution coordonnée des ressources ontologiques des ontologies de haut niveau. Pour qu'une ontologie puisse être intégrée dans la bibliothèque OBO, elle doit satisfaire les principes ontologiques suivants :

- Avoir un identifiant unique ;
- Définir textuellement tous ses termes ;
- Utiliser un langage formel commun : les ontologies doivent être décrites dans un langage d'ontologie standard ;
- Être orthogonale avec les autres ontologies d'OBO : les ontologies ne doivent pas se chevaucher pour profiter des bienfaits du développement modulaire et elles doivent être basées sur des ontologies de haut niveau, en l'occurrence BFO (Basic Formal Ontology) [Smith 2005b] et RO (Relation Ontology) [Smith 2005a] pour l'établissement des liens entre les concepts.
- Être ouverte : les ontologies doivent être libres d'utilisation ;
- Être bien documentée.

Comme elles sont développées autour de l'ontologie fondatrice BFO, les ontologies OBO distinguent les "continuants" et les "occurrents". Comme le montre la Figure 3.9, OBO comprend des ontologies à différents niveaux de granularité, du niveau de la molécule au niveau de l'organisme, en passant par le niveau de la cellule. Pour favoriser le développement collaboratif des ontologies biomédicales, un ensemble de développeurs d'ontologies OBO a initié la fonderie OBO (OBO foundry). Ce framework respecte les principes de base d'OBO ainsi que d'autres principes, à savoir : 1) le développement doit être fait en collaboration, 2) l'utilisation de relations définies sans ambiguïté, 3) la mise en place d'une procédure de traitement des remarques des utilisateurs et l'identification des versions successives des ontologies, 4) l'attribution d'un thème précis à l'ontologie.

La fonderie OBO a réussi à faire aligner plusieurs ontologies telles que l'ontologie GO (Gene Ontology) [Consortium 2008] et l'ontologie FMA (Foundational Model

RELATION TO TIME GRANULARITY	CONTINUANT			OCCURRENT	
	INDEPENDENT		DEPENDENT		
ORGAN AND ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FME, CPRO)	Phenotypic Quality (PaTO)	Organism-Level Process (GO)
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		Cellular Process (GO)
MOLECULE	Molecule (ChEBI, SO, RNAO, PRO)		Molecular Function (GO)		Molecular Process (GO)

FIGURE 3.9 – Les différents niveaux de granularité des ontologies OBO couvrant des données de type "BFO:continuant" et "BFO:occurrent"

of Anatomy) [Rosse 2003], des nouvelles ontologies ont été créées sur la base de ces principes et certaines ontologies ont été intégrées dans une seule ontologie, on peut donner l'exemple de l'ontologie CL (Cell Ontology) [Bard 2005] qui a été intégrée dans l'ontologie FMA pour la représentation des types de cellules.

### 3.5 Représentation de la sémantique des données d'imagerie avec des ontologies

Dans cette section, nous présentons les principales ontologies dans le domaine du traitement du contenu informationnel des images : des ontologies d'annotation, de marquage ou de bio-banques d'images.

#### 3.5.1 Le modèle Open Annotation

Le modèle OA (*Open Annotation*) [Sanderson 2013] est une infrastructure de description sémantique des annotations d'imagerie et des différentes relations entre elles. Ce modèle est conforme aux standards du W3C. Ses annotations sont conçues pour être à la fois expressives et simples afin de couvrir une large panoplie de cas d'utilisation du domaine de la bio-informatique. Le modèle OA résulte de la fusion des deux modèles d'annotation Annotation Ontology [Ciccarese 2011] et OA Collaboration [Bradshaw 2013].

Le modèle OA permet principalement d'annoter des observations d'identification, de classification et de description. Ces annotations font le lien entre des ressources de type "OA:body" et d'autres ressources de type "OA:target" qui sont référencées à travers des URLs. Les ressources "OA:body" décrivent les ressources de type "OA:cible" sous un format de texte, image, vidéo, etc. Elles peuvent aussi n'être associées qu'à une partie de la ressource de type "OA:target". Ainsi, l'infor-

Nom	type	Classes	Domaine	Format
SNOMED CT	système de codage	327128	santé	UMLS
MESH	système de codage	265414	santé	UMLS
LOINC	système de codage	192372	résultat de laboratoire	UMLS
NCIT	thésaurus	124824	recherche en cancérologie	UMLS
FMA	ontologie	104522	anatomie	OBO
OMIM	ontologie	93042	genetique	UMLS
Radlex	ontologie	46433	radiologie	-
ICD-10	système de classification	12445	pathologies	UMLS
HP	ontologie	15804	phénotypes	OBO
DO	ontologie	12479	pathologie	OBO
ATC	système de classification	6135	médicaments	-
OBI	ontologie	3071	expérimentations	OBO
CMO	ontologie	2738	mesures	OBO
PATO	ontologie	2541	phénotypes	OBO
UO	ontologie	381	unités de mesure	OBO
TMO	ontologie	225	études cliniques	-
IAO	ontologie	220	contenu informationnel	OBO
BSPO	ontologie	133	information spatial	OBO
OGMS	ontologie	124	médecine	OBO
CARO	ontologie	50	anatomie	OBO
BFO-2.0	ontologie	32	général	OBO

TABLE 3.1 – Listing de différentes ressources du Biportal NCBO ordonnées par le nombre des classes

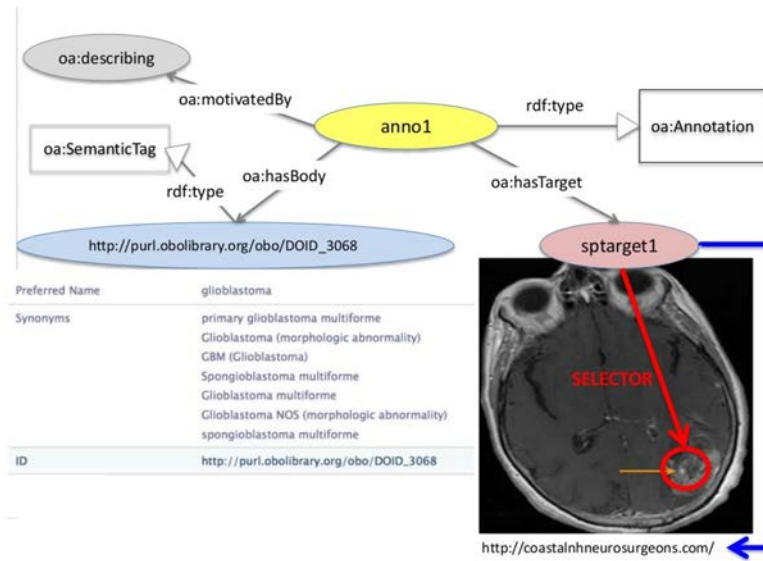


FIGURE 3.10 – Annotation avec le modèle OA d'une lésion de type glioblastome identifiée sur une séquence de type IRM [Ciccarese 2014]

mation d'extraction est décrite comme suit : une ressource nommée "OA:specific ressource" décrit le fragment de référence, une ressource nommée "OA:SELECTOR" décrit la méthode d'extraction et une ressource nommée "OA:Full ressource" fait référence à la ressource de type "OA:cible" d'origine. Pour illustrer, la Figure 3.10 montre un exemple d'annotation qui décrit une lésion qui est identifiée sur une séquence d'IRM. On note que toutes ces déclarations sont exprimées sous forme de triplets RDF.

Les nouvelles plates-formes d'annotation, telles que Domeo Annotation Web Toolkit, [Ciccarese 2012] ont démontré comment ces modèles peuvent supporter un ensemble très riche de fonctionnalités d'annotation.

### 3.5.2 L'ontologie OntoNeuroLOG

L'ontologie OntoNeuroLOG (ONL) [Gibaud 2011] permet de représenter des méta-données issues du domaine de la neuro-imagerie. Ces métadonnées concernent le contenu des images, les conditions d'acquisition d'image (par exemple, l'équipement de MR et les séquences de MR), les sujets impliqués, les méthodes d'acquisition et de traitement d'image (par exemple, correction ou segmentation). L'ontologie OntoNeuroLOG est basé sur l'ontologie fondamentale DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [Masolo 2003] et elle vise à fédérer des données hétérogènes de plusieurs bases de données distribuées dans le domaine de la neuro-imagerie.

Ce travail a donné lieu à la définition d'une ontologie des traitements d'images médicales (ONL-DP) qui représente les données d'imagerie ainsi que les processus d'imagerie. Les images médicales sont organisées selon trois axes sémantiques : la

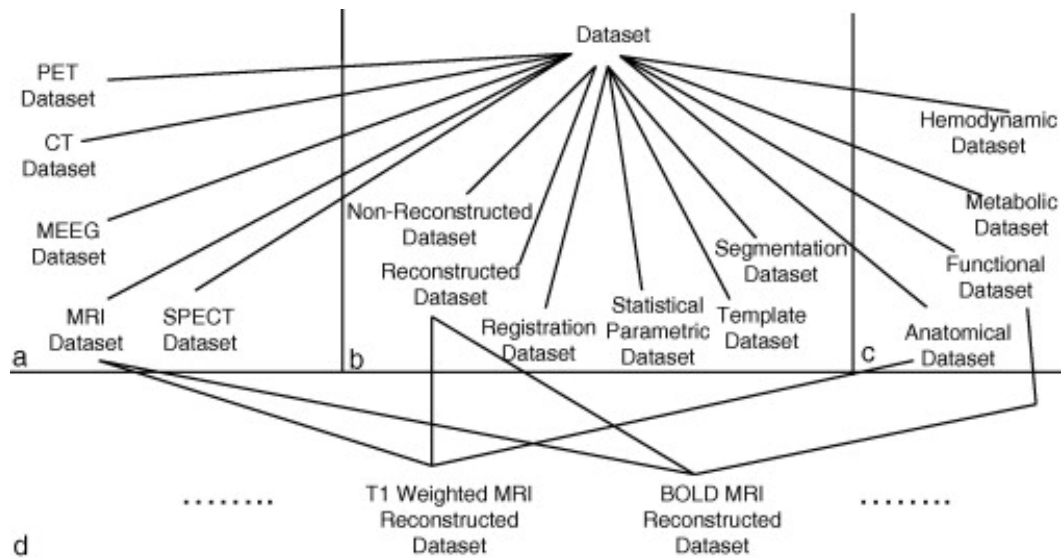


FIGURE 3.11 – Extrait des types de Datasets les plus généraux basés sur (a) la modalité, (b) le traitement, (c) l'entité explorée et (d) les types de Dataset de niveau inférieur qui héritent des propriétés des trois axes sémantiques précédents. [Temal 2008]

modalité d'image, le traitement d'image dont elles sont le résultat, l'entité explorée pendant l'examen. La Figure 3.11 illustre la taxonomie des images médicales, on note que le terme "Dataset" a été employé pour faire référence au contenu des images médicales concernant un sujet ou un groupe de sujets. ONL-DP définit les pré-traitements d'imagerie (le lissage, le recalage, le débruitage, etc.), les traitements d'imagerie (segmentation, etc.) et la quantification des paramètres d'imagerie (estimation du flux sanguin cérébral, calcul de l'anisotropie, etc.). ONL-DP est disponible sur le portail de la NCBO.

Une des limites de l'ontologie ONL est qu'elle utilise l'ontologie DOLCE et non pas BFO, ce qui restreint son emploi par d'autres ontologies biomédicales, ce qui va rendre difficile l'intégration des données (ultime objectif d'ONL).

### 3.5.3 L'ontologie QIBO

Buckler et al. ont proposé la première ontologie des biomarqueurs d'imagerie sous le nom de *Quantitative Imaging Biomarker Ontology* (QIBO) [Buckler 2013a] pour améliorer la gestion des données associées aux biomarqueurs d'imagerie de type quantitatif. QIBO vise principalement à faciliter l'accès et la recherche des données d'imagerie quantitative, à accélérer le développement des nouveaux biomarqueurs, et à fournir un framework pour réaliser des tests d'hypothèses, etc. Le but de QIBO est de "représenter, intégrer et harmoniser les connaissances hétérogènes du domaine des biomarqueurs d'imagerie".

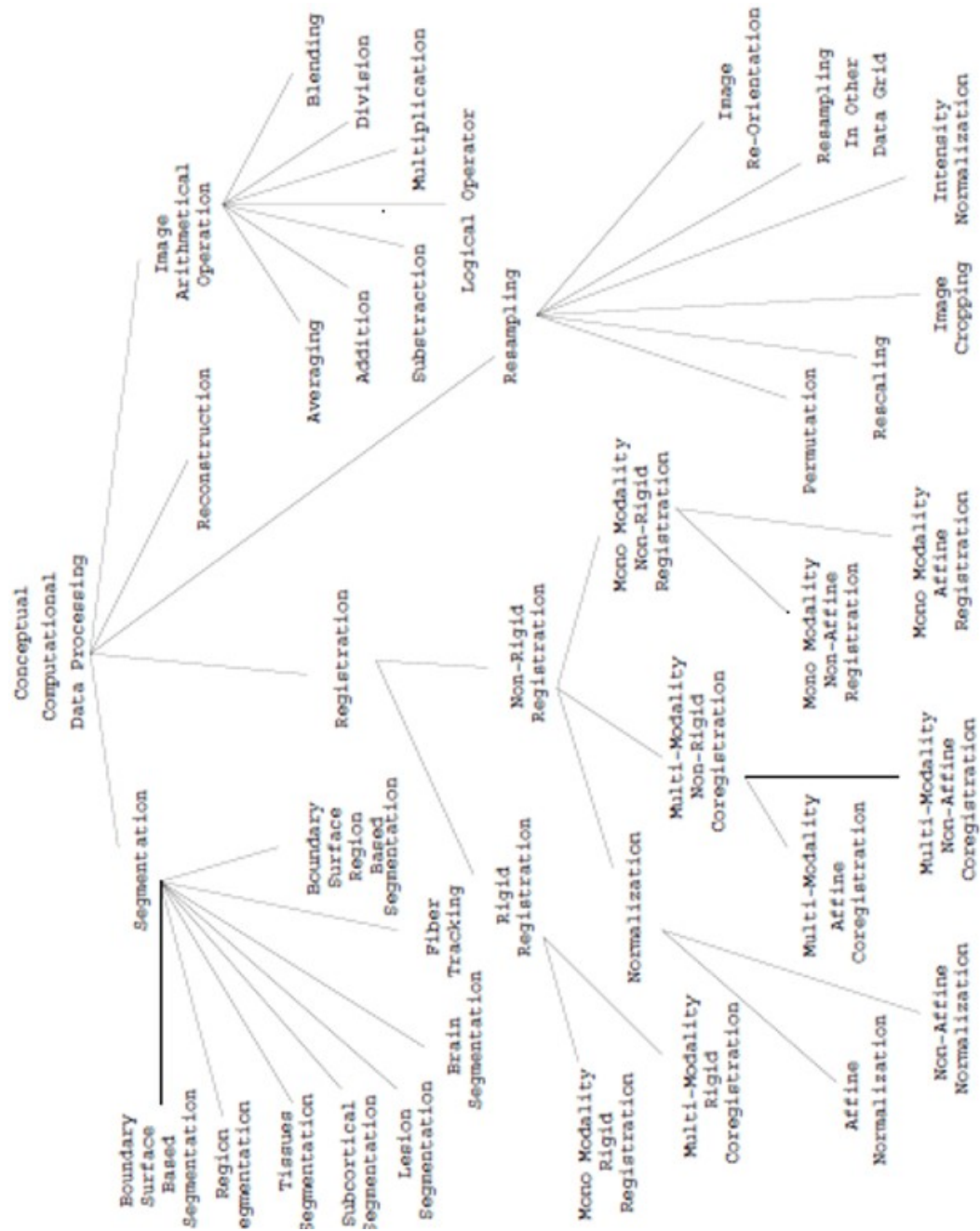


FIGURE 3.12 – Extrait de la taxonomie des traitements qui portent sur les "Data-sets", [Lynda 2008]



L'ontologie QIBO a introduit 488 classes, 56 propriétés d'objet et 72 propriétés de données pour représenter d'une manière sémantique les biomarqueurs d'imagerie. Comme le montre la figure 3.13, les principales classes de QIBO sont le sujet, le produit de contraste utilisé, l'organe anatomique, l'instrument de mesure, la méthode de mesure (algorithme), le processus biologique mis en question et le rôle du biomarqueur. Cette ontologie est la première initiative pour démêler les différentes entités intervenant dans le processus de mesurage mais elle est loin d'être la bonne solution pour résoudre le problème de la représentation des biomarqueurs d'imagerie vu :

- L'absence de définitions formelles et textuelles des termes proposés ;
- L'absence de quelques termes clés dans le domaine des biomarqueurs (la notion de région d'intérêt, validité du biomarqueur, etc.) ;
- La non réutilisation des ontologies déjà existantes et l'absence de référence à une ontologie fondamentale ;
- L'absence de définition formelle des relations entre les termes ;
- etc.

La modélisation du concept de biomarqueur d'imagerie dans QIBO ne prend pas en compte les aspects suivants :

- Séparation entre les deux axes sémantiques qualité mesurée et valeur de mesure.
- Confusion dans la description des paramètres quantitatifs et des paramètres qualitatifs. Par exemple la classe "QIBO:shape parameter" est définie comme un biomarqueur quantitatif tant dis qu'elle mesure plutôt un paramètre qualitatif qui est la forme "shape" ;
- Confusion entre la qualité biologique mesurée et le processus pathologique mis en jeu, de sorte qu'il n'est pas possible de chercher les biomarqueurs d'imagerie en utilisant la nature de la qualité mesurée.
- Absence de lien entre les images médicales et les valeurs de biomarqueurs qui sont dérivées à partir de ces dernières ;

Ce travail est néanmoins un point de départ intéressant pour permettre un raisonnement sur la sémantique des données des biomarqueurs, bien qu'il manque des définitions explicites et formelles des entités proposées. En outre, il n'utilise pas d'ontologie fondamentale, ce qui pose un problème majeur parce que le domaine des biomarqueurs implique de nombreux domaines complémentaires pour lesquels des ontologies existent déjà, mais nécessitent d'être intégrées.

#### **3.5.4 Le Système BiomRKRS**

*BiomRKRS* (Biomarker Retrieval and Knowledge Reasoning system) [Ofoghi 2014] est un système basé sur des ontologies et des vocabulaires existants et qui a pour objectif de faciliter le raisonnement sur les données associées au concept de biomarqueur d'une façon générale. Le système BiomRKRS est composé de trois composantes logicielles : (1) l'ontologie de BiomRKRS et le système de gestion qui lui est associé, (2) la base de données du système et (3) le système de

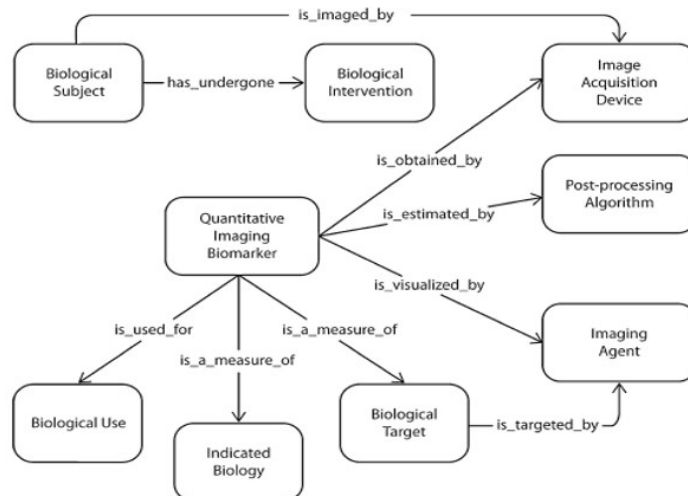


FIGURE 3.13 – Structure générale de l'ontologie QIBO, extraite de [Buckler 2013a]

raisonnement de BiomRKRS.

### 3.5.4.1 L'ontologie BiomRKRS

Comme le montre la Figure 3.14 les principaux concepts de l'ontologie BiomRKRS sont : la maladie associée au biomarqueur, l'état de validation (pour préciser l'état de qualification du biomarqueur), le rôle clinique du biomarqueur, la technique de mesurage du biomarqueur, les preuves scientifiques, les critères de stratification, et d'autres concepts qui sont détaillés dans l'article. Contrairement à QIBO, BiomRKRS a utilisé des ontologies et des vocabulaires existants pour décrire ses concepts, notamment : Gene Ontology (GO) pour la description des processus biologiques mis en jeu, Experimental Factor Ontology (EFO) [Malone 2010] pour la description de la stratification, ICD (International Classification of Diseases) 10 pour la description de la maladie, LOINC pour la description de la technique de mesurage.

L'ontologie BiomRKRS est décrite en OWL/XML et elle est maintenue par un système qui permet d'assurer l'intégration des sources externes ainsi que la prise en considération des mises à jour. On note qu'un convertisseur interne a été développé pour la génération d'un format OWL de quelques terminologies (par exemple, ICD-10).

Techniquement, l'ontologie BiomRKRS a quelques limites et jusqu'à maintenant, l'ontologie n'est pas accessible de façon libre. De plus, l'ontologie BiomRKRS n'est pas basée sur une ontologie de haut niveau pour la représentation de ses concepts et relations. Les auteurs de ce travail n'ont pas cherché à aligner les relations de l'ontologie à RO ; des relations très spécifiques ont été définies (par exemple "IsEvidenceOf", "IsGivenTo", etc.). Enfin, l'ontologie utilise aussi des ressources externes qui ne sont pas toutes des ontologies, ce qui réduit son expressivité sémantique.

Du point de vue de la modélisation, l'ontologie n'a pas introduit plusieurs no-

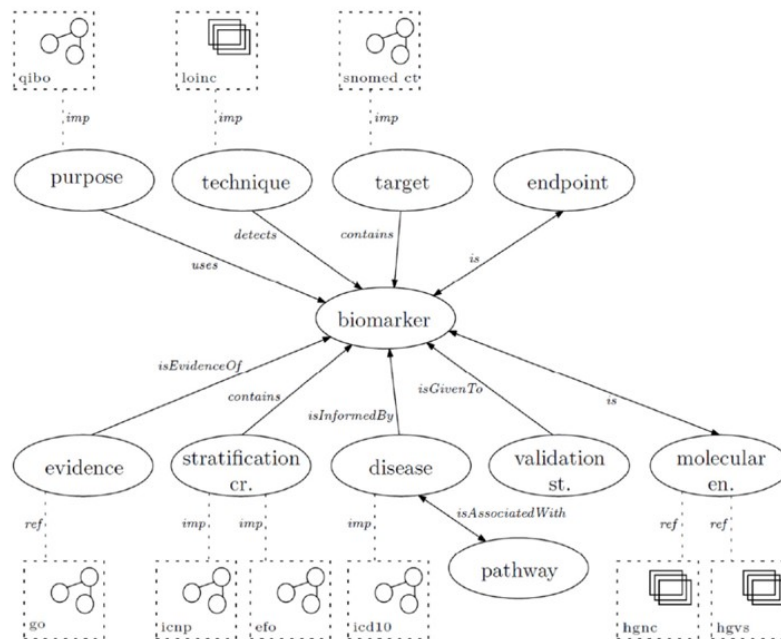


FIGURE 3.14 – Les principaux concepts de l’ontologie BiomRKRS, extraite de [Ofoghi 2014]

tions importantes liées à la définition des biomarqueurs notamment : la notion de protocole et de processus de mesure, les applications des biomarqueurs, la valeur de biomarqueur, et la catégorisation de ce dernier (qualitatif ou quantitatif).

### 3.5.4.2 Le corpus de données du système BiomRKRS

Cette base de données contient principalement les données associées aux instances des biomarqueurs ainsi que les informations sur l’emplacement des modules externes de l’ontologie BiomRKRS. Actuellement, les instances de biomarqueurs qui sont incluses dans la base sont extraites manuellement par des experts du domaine. Les sources d’information employées sont les publications scientifiques sur deux pathologies spécifiques qui sont : la polyarthrite rhumatoïde et l’ostéoporose. Nous pensons que le système BiomRKRS doit être validé sur un corpus de données plus grand pour mieux crédibiliser le travail, d’autant plus que les auteurs ont choisi de représenter tous les types de biomarqueurs du domaine biomédical.

### 3.5.4.3 Le système de raisonnement de BiomRKRS

Le système de raisonnement de BiomRKRS est un système à base de connaissances qui permet de répondre aux requêtes des utilisateurs en utilisant des inférences sémantiques basées sur RDFS. Ainsi, le système peut assister les utilisateurs dans :

- la recherche de la liste de biomarqueurs d'une certaine pathologie en fonction de leurs rôles cliniques et en fonction des gènes responsables de la maladie ;
- la sélection des travaux scientifiques qui analysent l'importance de certains gènes associés à un mécanisme pathologique ;
- l'affichage des techniques de mesurage d'un biomarqueur à appliquer dans un contexte donné.

### 3.5.5 Une représentation des biomarqueurs fondée sur l'ontologie OGMS

Dans leur étude sur les biomarqueurs [Ceusters 2015], Ceusters et al. se sont basés sur les concepts de l'ontologie OGMS (Ontology for General Medical Science) qui sont listés dans la figure 3.16. Dans ce travail, le concept de biomarqueur est défini comme une caractéristique qui est observable et évaluable, c'est-à-dire une caractéristique qui est toujours identifiable dans un processus d'observation et d'évaluation. Ainsi, les auteurs ont défini le biomarqueur comme un "OGMS:bodily feature" qui subsume trois catégories de biomarqueurs qui sont disjointes (voir Figure 3.15) : la première catégorie est le "Material biomarker" qui fait référence à la structure anatomique "OGMS:bodily component" qui est l'objet de l'observation, la deuxième catégorie est le "Quality biomarker" qui décrit la qualité associée à la structure anatomique observée (modélisée comme un "OGMS:bodily quality") et la troisième catégorie est le "Process biomarker" qui évalue si le processus réalisé est normal ou pathologique (modélisé comme un "OGMS:bodily process"). Par conséquent, d'après Ceusters et al. la définition semi-formelle du concept de biomarqueur est la suivante : Biomarker = def. "Material Biomarker", "Quality Biomarker" ou "Process Biomarker".

La première limite de cette proposition est qu'elle a défini les biomarqueurs comme des entités qui sont observées dans le corps de l'être humain et qu'elle n'a pas inclus l'aspect de processus de mesure. Cette exclusion de l'aspect de mesurage n'a pas été justifiée bien qu'il s'agisse d'un aspect important des biomarqueurs d'imagerie (voir les travaux du groupe QIBA). La deuxième limite est que ce travail n'est pas facilement extensible pour la représentation des biomarqueurs dans le domaine médical. Nous pensons que QIBO et BiomRKRS sont préférables à cette solution.

### 3.5.6 Le projet THESEUS MEDICO

Le projet allemand THESEUS MEDICO [Seifert 2010] avait pour objectif de créer des moteurs de recherche intelligents pour faciliter l'exécution des requêtes relatives aux images médicales et aux comptes rendus radiologiques afin de garantir un bon usage des informations biomédicales. Le projet THESEUS MEDICO est basé sur l'emploi des technologies du Web sémantique. L'ontologie MEDICO utilise l'ontologie FMA [Rosse 2003] pour la description des structures anatomiques, Radlex pour la description des observations radiologiques (extraction du sous arbre

	T	Definition
1	C	<b>BIOMARKER</b> =def.– <b>MATERIAL BIOMARKER</b> , <b>QUALITY BIOMARKER</b> or <b>PROCESS BIOMARKER</b> .
1a	U	<b>MATERIAL BIOMARKER</b> =def.– BODILY COMPONENT <i>instances of</i> which are <i>capable of</i> being assessed objectively to determine what kind of PROCESSES either they or some QUALITIES that <i>depend on</i> them <i>result from</i> .
1b	U	<b>QUALITY BIOMARKER</b> =def.– BODILY QUALITY <i>inhering in</i> a BODILY COMPONENT <i>c</i> and in virtue of which <i>c</i> is an <i>instance of</i> <b>MATERIAL BIOMARKER</b>
1c	U	<b>PROCESS BIOMARKER</b> =def.– BODILY PROCESS <i>p</i> <i>instances of</i> which are <i>capable of</i> being assessed objectively to determine whether <i>p</i> is a <i>realization of</i> NORMAL or of ABNORMAL HOMEOSTASIS.
2	C	<b>DISEASE BIOMARKER FOR DISEASE OF TYPE X</b> =def.– <b>BIOMARKER</b> instances of which either start to exist or become clinically abnormal during a <i>part</i> of the DISEASE COURSE which <i>realizes</i> the DISEASE OF TYPE X
3	C	<b>DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> =def.– <b>MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> , <b>QUALITY DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> , or <b>PROCESS DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b>
3a	U	<b>MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> =def.– <b>MATERIAL BIOMARKER</b> <i>instances of</i> which are <i>capable of</i> being assessed objectively to determine whether a DISEASE OF TYPE X <i>inheres in</i> the ORGANISM of which these instances <i>are part</i> .
3b	U	<b>QUALITY DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> =def.– BODILY QUALITY which <i>inheres in</i> a BODILY COMPONENT <i>c</i> and in virtue of which <i>c</i> is an <i>instance of</i> <b>MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b>
3c	U	<b>PROCESS DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> =def.– BODILY PROCESS in which a BODILY COMPONENT <i>c</i> <i>participates</i> and in virtue of which <i>c</i> is an <i>instance of</i> <b>MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> .
4	U	<b>PROGNOSTIC BIOMARKER FOR DISEASE OF TYPE X</b> =def.– <b>DISEASE BIOMARKER FOR DISEASE OF TYPE X</b> which starts to exist during a <i>part p</i> of a DISEASE COURSE <i>d</i> which <i>realizes</i> the DISEASE OF TYPE X and where <i>p</i> is such that it <i>regulates</i> the further unfolding of <i>d</i> .

**Legend:** T = defined class (C) or universal (U). **BOLD SMALL CAPS** = generic entities defined in this table. **SMALL CAPS** = generic entities defined in other BFO-based ontologies; *bold italics* = relations.

FIGURE 3.15 – Définitions semi-formelles des principaux concepts du biomarqueur, extraite de [Ceusters 2015]

<b>DISORDER</b> =def. – causally relatively isolated combination of BODILY COMPONENTS that is (a) clinically abnormal and (b) maximal, in the sense that it is not a <i>proper part of</i> some larger such combination.
<b>PATHOLOGICAL PROCESS</b> =def. – BODILY PROCESS that is a <i>manifestation of</i> a DISORDER.
<b>DISEASE</b> =def. – DISPOSITION (i) to undergo PATHOLOGICAL PROCESSES that (ii) <i>inheres in</i> an ORGANISM because of one or more DISORDERS <i>in</i> that ORGANISM.
<b>DISEASE COURSE</b> =def. – The totality of all PROCESSES through which a given DISEASE instance <i>is realized</i> .
<b>MANIFESTATION OF A DISEASE</b> =def. – BODILY FEATURE of an ORGANISM that is (a) a deviation from clinical normality that exists in virtue of the <i>realization of</i> a DISEASE and (b) is observable.
<b>ABNORMAL HOMEOSTASIS</b> =def. – HOMEOSTASIS that is clinically abnormal for an ORGANISM of a given type and age in a given environment.
<b>NORMAL HOMEOSTASIS</b> =def. – Homeostasis of a type that is not clinically abnormal.
<b>BODILY COMPONENT</b> =def. – MATERIAL ENTITY <i>within or on</i> the surface of the body of an ORGANISM.
<b>BODILY PROCESS</b> =def. – PROCESS unfolding <i>in or on</i> the body in which BODILY COMPONENTS <i>participate</i>
<b>BODILY QUALITY</b> =def. – QUALITY which <i>inheres in</i> a BODILY COMPONENT.
<b>BODILY FEATURE</b> =def. – BODILY COMPONENT, BODILY QUALITY or BODILY PROCESS.

**Legend:** SMALL CAPS = generic entities defined in this table or in BFO; *bold italics* = relations

FIGURE 3.16 – Les principaux concepts OGMS utilisés, extraite de [Ceusters 2015]

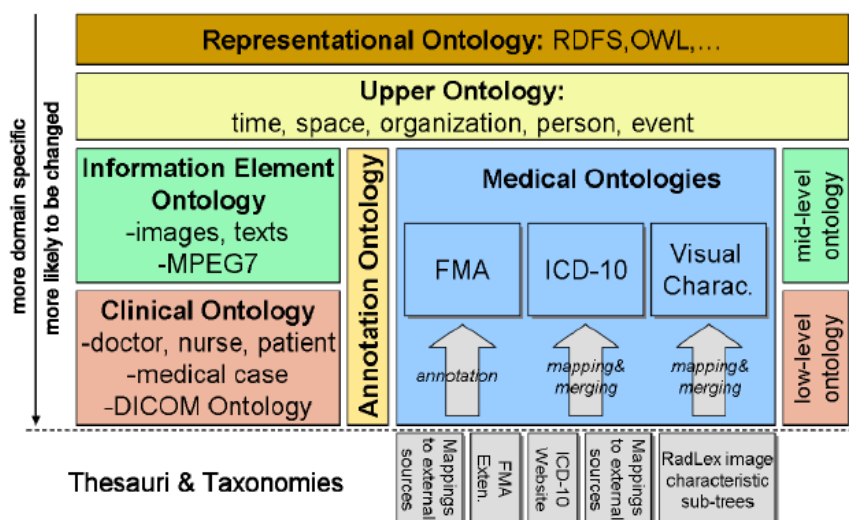


FIGURE 3.17 – Structure de l'ontologie MEDICO [Seifert 2010]

du concept "modifier and imaging observation characteristic") et ICD 10 pour la description des pathologies. Comme le montre la Figure 3.17, d'autres ontologies sont employées pour la description du contexte médical (patient, médecin, etc.) et le "mapping" entre les concepts de l'ontologie de haut niveau et de l'ontologie de domaine (appelée Information Element Ontology).

Un des axes de développement du projet est la représentation sémantique des mesures radiologiques ; le modèle MCI (Model for Clinical Information) décrit les différentes observations radiologiques, et entre autres, les mesures (voir Figure 3.18 qui illustre les différents concepts qui sont inclus dans la représentation de la taille des ganglions avec le modèle MCI) [Oberkampff 2016]. L'enrichissement sémantique des observations radiologiques a permis d'assister les cliniciens dans l'évaluation de la normalité des mesures. Les valeurs anormales de biomarqueurs d'imagerie sont mises en couleur rouge dans les rapports médicaux des patients.

Toujours dans le contexte du projet THESEUS-MEDICO, l'ontologie FMA a été étendue dans ce travail [Seifert 2010] pour décrire la dimension spatiale de l'emplacement des organes des patients pour la prise en charge du lymphome. L'emploi de FMA a permis d'améliorer les algorithmes de segmentation automatique des structures anatomiques sur les images CT et IRM via la description sémantique du contenu de l'image.

Dans un autre travail [Oberkampff 2012], le modèle MCI a été employé pour expliciter d'une manière formelle les relations entre les pathologies et leurs symptômes afin d'automatiser le classement des pathologies probables et la planification des prochains examens médicaux. Ce travail est le fruit de l'intégration des connaissances du modèle MCI avec celles de l'ontologie HDO (Human Disease Ontology)<sup>14</sup> qui décrit environ 8000 pathologies en se basant sur des lexiques tels que :

14. <http://disease-ontology.org/>



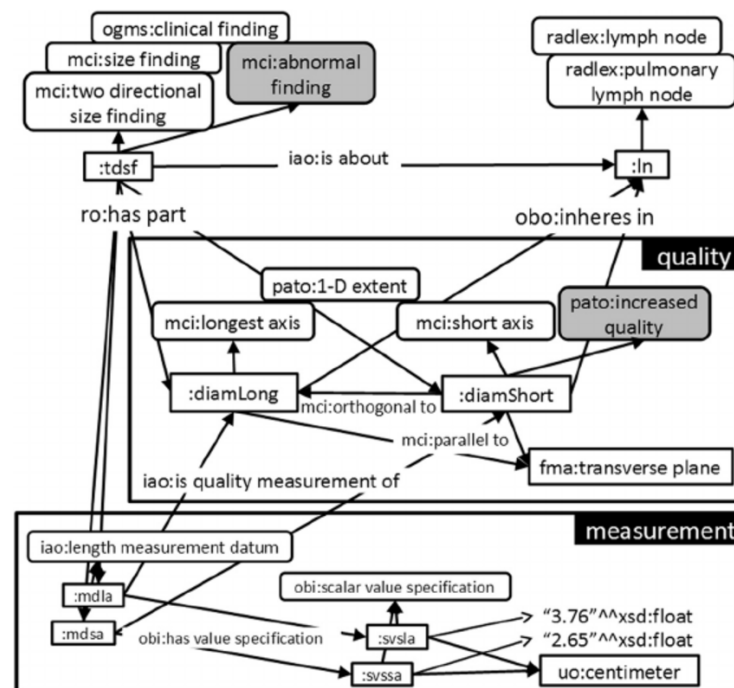


FIGURE 3.18 – Représentation de la taille des ganglions lymphatiques dans le plan axial selon les critères RECIST avec le modèle MCI [Oberkampff 2012]

MSH, SNOMED, UMLS, ICD, etc. et 15 relations de type "doid:has\_symptom" , "doid:transmitted\_by", etc. (par exemple : "infection has\_symptom fever, infection has\_symptom headache").

## 3.6 Discussion

### 3.6.1 Problématique générale des contributions de la thèse

La médecine de demain exige l'intégration des résultats de quantification dans les comptes rendus radiologiques pour aider à compléter le diagnostic radiologique traditionnel et améliorer sa précision. Ainsi, le compte rendu radiologique de demain doit être personnalisé, standardisé et structuré. Jusqu'à aujourd'hui les comptes rendus radiologiques sont encore des documents narratifs qui ne prennent pas en compte des données quantitatives d'imagerie et qui sont ultérieurement difficiles à ré-exploiter. Ces exigences soulignent le besoin de définir d'une manière explicite et formelle des définitions des informations de provenance relatives au concept de biomarqueur d'imagerie qui est au cœur de la médecine de précision. Ces informations peuvent être classées en deux grandes catégories : (1) les données acquises, à savoir les données brutes et les conditions d'acquisition aussi bien que (2) les données traitées et interprétées, à savoir les données résultantes de l'exécution d'une procédure spécifique.

Tous les travaux existants autour de la représentation sémantique des biomarqueurs sont intéressants et complémentaires, mais ils ne fournissent pas une solution qui est prête à l'emploi. L'ontologie la plus mature est probablement BiomRKRS, mais elle n'est pas accessible au public, et elle ne couvre pas tous les aspects du domaine des biomarqueurs d'imagerie. QIBO doit être alignée à d'autres ontologies biomédicales pour enrichir ses classes et relations. Enfin, les projets QIBA constituent des contributions intéressantes, notamment du fait de leur focalisation sur la notion de protocole. Cependant, ces informations ne sont pas disponibles sous un format ontologique qui facilite leur adoption dans les processus d'annotation des images. Le standard DICOM doit être enrichi avec de nouveaux modes d'implémentation des comptes rendus structurés des procédures d'imagerie, notamment destinés à leur gestion dans les futures biobanques d'images.

Dans le cadre de la thèse nous nous sommes particulièrement intéressés à l'étude de deux sujets : (1) la conception d'une représentation sémantique générale des biomarqueurs d'imagerie (2) la description et l'expérimentation de la valeur ajoutée de la formalisation des descripteurs d'imagerie via une approche réaliste, complémentaire du premier travail. Ces deux travaux de la thèse ont pour objectif de répondre aux besoins de la communauté de l'imagerie médicale quantitative à mieux intégrer, formaliser et réutiliser ces informations (inter et intra sites). Nous avons opté dans ces deux travaux par l'emploi des technologies du Web sémantique pour répondre à ces besoins.

### 3.6.2 Contributions de la thèse

**Première contribution de la thèse :** L'objectif de ce travail est de créer une ontologie générique IBO pour la formalisation de la représentation des biomarqueurs d'imagerie. La valeur ajoutée de cette ontologie découle notamment de son caractère générique, c'est à dire de son aptitude à couvrir des besoins médicaux ou de recherches diversifiés, notamment en terme de modalités d'imagerie, de spécialités médicales, de domaines pathologiques visés, et de problématiques de recherche clinique et translationnelle. La représentation des données de provenance des biomarqueurs d'imagerie fait intervenir plusieurs éléments d'information qui proviennent de domaines différents : les données brutes et dérivées d'imagerie, les traitements en imagerie, les phénomènes biologiques, les résultats et protocoles de mesure, les données cliniques, les rapports structurés, les phénotypes d'imagerie, les processus décisionnels, etc. Un des défis de ce travail est d'intégrer ces différentes sources d'information (ontologies existantes et spécialisées) pour correctement couvrir les différents aspects du domaine des biomarqueurs d'imagerie.

Ce travail vise à faciliter le partage à grande échelle des informations relatives aux biomarqueurs d'imagerie afin de faciliter le développement de nouveaux outils de traitement des images médicales, en facilitant la réalisation d'études visant à comparer les résultats obtenus avec des méthodes existantes. Ce partage doit également faciliter l'application de méthodes de traitement existantes à des nouveaux contextes cliniques.



**Deuxième contribution de la thèse :** L'objectif de ce travail est de montrer l'intérêt de la représentation sémantique des biomarqueurs d'imagerie dans un domaine concret. Nous avons choisi d'annoter les critères d'imagerie associés aux images IRM du glioblastome multiforme (GBM) et qui sont présents dans la terminologie VASARI. Ici, nous avons conçu un modèle ontologique décrivant les entités mises en jeu dans l'évaluation des tumeurs gliales pour aller au-delà de la description "simpliste" proposée par l'ontologie IBO. En d'autres termes, le modèle développé permet d'étendre la description informationnelle fournie par IBO afin de mieux décrire la réalité biologique observée chez le patient.

Ce travail vise à enrichir les résultats d'imagerie des systèmes d'analyse d'imagerie via les connaissances du modèle sémantique sur, par exemple, la composition des lésions, volumes, etc. Cet axe de la thèse a donné lieu à un travail d'expérimentation sur le corpus de données REMBRANDT qui est issu du TCIA américain (The Cancer Imaging Archive). Cette base contient des observations VASARI relatives à 34 patients qui sont produites par trois spécialistes.



## Chapitre 4

# Vers une Représentation Sémantique des Biomarqueurs d’Imagerie

### Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>91</b>
<b>4.2</b>	<b>Contexte et objectif du travail</b>	<b>92</b>
<b>4.3</b>	<b>Matériel et méthode</b>	<b>93</b>
4.3.1	Portée de l’ontologie des biomarqueurs d’imagerie	93
4.3.2	Intégration des ontologies	94
4.3.3	Réutilisation des ontologies en fonction des contextes	99
4.3.4	Construction des modules ontologiques	111
<b>4.4</b>	<b>Résultat</b>	<b>112</b>
4.4.1	Développement de l’ontologie IBO	112
4.4.2	Application de l’ontologie IBO pour la représentation de quelques biomarqueurs d’imagerie	117
<b>4.5</b>	<b>Discussion</b>	<b>120</b>

---

### 4.1 Introduction

Dans le chapitre précédent nous avons présenté des travaux existants dans le domaine de la modélisation et de l’annotation des données d’imagerie médicale. Dans ce chapitre, nous présentons notre modèle sémantique de modélisation des biomarqueurs d’imagerie. Nous expliquons dans la Section 5.3 la problématique étudiée et les motivations de notre travail. Dans la Section 5.4, nous détaillons la portée de l’ontologie pour mieux cerner le domaine à représenter, présentons la méthodologie de développement de l’ontologie IBO (Imaging Biomarker Ontology) et décrivons les étapes d’intégration, de réutilisation et d’extraction des modules ontologiques d’IBO.

La Section 4.4.2 présente quelques exemples d’utilisation de l’ontologie IBO pour la représentation de quelques exemples de biomarqueurs d’imagerie. Nous consacrons la Section 4.5 à la discussion des apports de l’ontologie IBO ainsi que les points à améliorer dans notre travail.

## 4.2 Contexte et objectif du travail

Comme il a été souligné dans le Chapitre 2, il est devenu important d’être en mesure de partager des informations sur les biomarqueurs d’imagerie pour garantir la compréhension, le partage et l’intégration des informations qui leur sont associées. Or, ce concept n’est pas facile à cerner, car il fait intervenir différentes notions : la qualité biologique estimée par le biomarqueur d’imagerie en question, le processus de mesure de la valeur du biomarqueur d’imagerie et le rôle du biomarqueur d’imagerie dans un processus décisionnel. Ainsi, en 2010 le "National Institute of Health" américain a souligné la nécessité de définir d’une manière cohérente et précise le vocabulaire lié aux biomarqueurs [Ball 2010] : "the committee observed a great deal of inconsistent and imprecise definition and use of terms relevant to biomarkers and biomarker evaluation. Consistent, precise definition and use of terms is critical for biomarker evaluation because it is a topic important across disciplines and has been for several decades."

En effet, aucun des travaux existants dans la littérature (état de l’art détaillé dans le Chapitre 3) n’a réussi à définir un vocabulaire qui couvre ce domaine de façon cohérente. Par exemple, dans certains travaux, le terme biomarqueur d’imagerie est employé pour faire référence à une mesure d’une qualité biologique chez un patient  $x$  (taille de la tumeur, concentration du produit de contraste dans une région d’intérêt, etc.) [Ceusters 2015]. Dans d’autres travaux scientifiques [VIM 2012], ce même terme est utilisé pour désigner non pas une mesure d’une qualité chez un patient, mais plutôt un outil de mesure qui assure la reproductibilité et la précision du processus de mesure de cette qualité biologique (QIBA). L’objectif de cet outil de mesure est de minimiser la variabilité relatives aux appareils de mesure, patients et temps. Ce concept de biomarqueur d’imagerie est aussi considéré chez certains scientifiques comme un couple (qualité biologique, degré de validation) qui a été validé par un des organismes (e.g. FDA) pour un usage particulier [ESR 2010], par exemple diagnostique, prédictif ou d’évaluation de la réponse à un traitement.

Pour toutes ces raisons et aussi parce que ce concept de biomarqueur d’imagerie promis à une utilisation à grande échelle, nous proposons ici de définir de façon cohérente le vocabulaire qui lui est associé. Ceci est réalisé sous la forme d’une ontologie de haut niveau des biomarqueurs d’imagerie qui prend en considération les trois principaux aspects cités : la qualité biologique, l’outil d’aide à la décision et l’outil de mesure. Ainsi, l’objectif principal de notre travail est de modéliser le concept de biomarqueur d’imagerie, en articulant convenablement les trois axes sémantiques qui sont : la qualité biologique mesurée, l’outil de mesure utilisé et le rôle que le biomarqueur d’imagerie joue dans le processus décisionnel pour lequel il a été va-

lidé. Pour cela, nous proposons une ontologie générique, appelée Imaging Biomarker Ontology (IBO), prenant en considération les travaux antérieurs mentionnés dans le Chapitre 2 et les intégrant dans un ensemble sémantiquement cohérent.

## 4.3 Matériel et méthode

### 4.3.1 Portée de l'ontologie des biomarqueurs d'imagerie

En termes de modélisation de l'information, la représentation des biomarqueurs d'imagerie peut être envisagée sous trois aspects complémentaires :

Le premier aspect concerne le mesurage de la valeur de biomarqueur d'imagerie par un médecin spécialiste afin de répondre à une demande diagnostique ou thérapeutique de la part d'un clinicien. Ceci nécessite la modélisation de l'organisation en étapes de cette activité qui suit un protocole bien défini au préalable. On peut citer quelques étapes d'organisation d'un processus diagnostique ou thérapeutique :

- l'acquisition des images (par exemple images servant au diagnostic, ou images pré-opératoires dans le contexte chirurgical) ;
- le pré-traitement des images pour répondre au mieux à la question clinique ;
- l'analyse des images et l'extraction des mesures pour étudier la pathologie ;
- la rédaction du compte rendu radiologique et la sélection des images les plus représentatives (en imagerie diagnostique) ;
- la préparation du geste chirurgical, la mise en oeuvre du plan d'intervention retenu et la documentation du déroulement de l'opération dans le contexte de l'imagerie thérapeutique.

Le second aspect concerne le critère d'imagerie lui-même et il souligne le besoin de la bonne spécification de la nature de la mesure estimée chez le patient. Ceci doit inclure principalement les informations suivantes : les séquences d'images simples et combinées utilisées, la région d'intérêt associée à la mesure, la valeur de la mesure, l'unité de mesure, la nature de la mesure, la localisation de la structure anatomique/pathologique étudiée, etc. (ces paramètres sont résumés dans la Figure 4.2). Dans les comptes rendus DICOM SR, les principales mesures qui doivent être supportées sont : les mesures linéaires (largeur, diamètre, etc.), les mesures volumétriques, les mesures calculées à partir de régions d'intérêt, les mesures dérivées d'autres mesures, et les évaluations qualitatives (voir Figure 4.3). Ces mesures sont catégorisées de la façon suivante :

- Mesures simples comme par exemple, les segments, les segments biorthogonaux, segments triorthogonaux ou les angles.
- Mesures calculées à partir de régions d'intérêt : le coefficient d'atténuation (en Hounsfield), les valeurs SUVs (en g/ml ou cm<sup>2</sup>/ml), etc. ;
- Mesures dérivées d'autres mesures (la somme des plus longs diamètres (en mm), la somme du produit entre "short diameter" et "long diameter" (en mm<sup>2</sup>), la déviation par rapport à la "baseline" de la somme des plus longs diamètres (en pourcentage), etc.)

- Annotations textuelles (entre autres les évaluations qualitatives de la méthode RECIST).

Le troisième aspect qui doit être pris en compte dans cette modélisation concerne le rôle du critère objectif étudié et son degré de validation. L'accès à ces informations est un élément important pour faciliter la comparaison des examens du patients et sa prise en charge.

Notre modèle vise à couvrir la représentation des biomarqueurs d'imagerie de type qualitatif et quantitatif : les biomarqueurs d'imagerie quantitatifs correspondent à ce que le Vocabulaire International de la Métrologie (VIM) définit comme une mesure et les biomarqueurs d'imagerie qualitatifs correspondent à des évaluations souvent désignées sous la forme de scores. Pour résumer, dans la vision d'IBO un biomarqueur d'imagerie est un concept qui doit être caractérisé d'une manière simplifiée de la façon suivante : un biomarqueur d'imagerie est le **résultat d'une analyse quantitative ou qualitative** d'un ou plusieurs paramètre(s) d'imagerie ET qui joue un **rôle clinique** particulier dans un ou plusieurs processus clinique(s).

L'ontologie IBO fait intervenir des termes qui proviennent de domaines différents (imagerie médicale, biologie, traitement de l'image, méthodes de mesurage, recherche clinique, etc.). Nous avons aussi utilisé une ontologie fondamentale pour assurer une cohérence sémantique au sein d'IBO.

### 4.3.2 Intégration des ontologies

Pour intégrer des classes de différentes ontologies, nous avons choisi de développer IBO autour de l'ontologie fondatrice BFO (Basic Formal Ontology). L'ontologie BFO <sup>1</sup> [Smith 2005b] est une ontologie de haut niveau qui fournit un cadre de modélisation commun pour faciliter l'intégration des données issues d'ontologies de domaines divers. Dans la bibliothèque OBO, l'ontologie BFO est l'épine dorsale des ontologies répertoriées ainsi elle constitue le socle intégrateur des différents modules d'ontologies. Les concepts de BFO peuvent se répartir en trois grandes catégories [Bittner 2004] [Bittner 2003] [Grenon 2004], ici nous ne présentons que deux catégories :

- Les entités nommées "**BFO:continuant**" et les processus nommés "**BFO:occurrent**" : la catégorie "**BFO:continuant**" modélise les entités qui persistent à travers le temps (organisme, atome, température, etc.). La catégorie "**BFO:occurrent**" modélise les événements auxquels les "**BFO:continuant**" participent (déplacement d'un objet physique, processus de diagnostic médicale, processus de mesurage de la température, etc.).
- Les entités nommées "**BFO:independent continuant**", "**BFO:generically dependent continuant**" et "**BFO:specifically dependent continuant**" : la catégorie "**BFO:continuant**" est répartie en entités dites "indépendantes" et d'autres "dépendantes". Par exemple, une structure anatomique est indépendante tandis que la qualité physique dépend du porteur (par exemple,

---

1. <http://www.ifomis.org/bfo>

TID 1410								
Planar ROI Measurements								
Type: Extensible      Order: Significant								
	NL	Rel with Parent	VT	Concept Name	VM	Req Type	Condition	Value Set Constraint
1			CONTAINER	EV (125007, DCM, "Measurement Group")	1	M		
2	>	HAS OBS CONTEXT	TEXT	DT(112039, DCM, "Tracking Identifier")	1	M		
3	>	HAS OBS CONTEXT	UIDREF	EV (112040, DCM, "Tracking Unique Identifier")	1	M		
4	>	HAS OBS CONTEXT	SCoord	EV (111030, DCM, "Image Region")	1	MC	XOR Row 6	GRAPHIC TYPE = not {MULTIPOINT}
5	>>	SELECTED FROM	IMAGE		1	M		
6	>	HAS OBS CONTEXT	IMAGE	EV (121214, DCM, "Referenced Segmentation Frame")	1	MC	XOR Row 4	Reference shall be to a Segmentation Image, with a single value specified in Referenced Frame Number, and with a single value specified in Referenced Segment Number
7	>	HAS PROPERTIES	IMAGE	EV (121233, DCM, "Source image for segmentation")	1	MC	IFF Row 6	
8	>	CONTAINS	IMAGE	EV (121200, DCM, "Illustration of ROI")	1	U		
9	>	CONTAINS	INCLUDE	DTID (1419) ROI Measurements	1-n	M		\$Measurement = \$Measurement \$Units = \$Units \$ModType = \$ModType \$ModValue = \$ModValue \$Method = \$Method \$Derivation = \$Derivation \$TargetSite = \$TargetSite \$TargetSiteMod = \$TargetSiteMod \$Equation = \$Equation \$RefAuthority = \$RefAuthority \$RangeAuthority = \$RangeAuthority \$DerivationParameter = \$DerivationParameter \$DerivationParameterUnits = \$DerivationParameterUnits

FIGURE 4.1 – Le template "TID 1410" nommé "Planar ROI measurements" structure les informations relatives aux mesures qui sont calculées à partir des régions d'intérêt planaires. Ces ROIs peuvent être référencées sur l'image à travers des formes de type "SCoord" ou bien des images de segmentation.

\$Measurement	Coded term or Context Group for Concept Name of measurement
\$Units	Units for the measurement
\$ModType	Modifier Name for Concept Name of measurement
\$ModValue	Modifier Value for Concept Name of measurement
\$Method	Value for Measurement Method
\$Derivation	Value for Measurement Derivation
\$TargetSite	Value for Anatomic Location of measurement
\$TargetSiteMod	ModifierValue for Anatomic Location of measurement
\$Equation	Coded term or Context Group for the equation or table from which the measurement was derived or computed
\$RefAuthority	Bibliographic reference or authority for statistical properties of a reference population
\$RangeAuthority	Bibliographic reference or authority for the normal range of the measurement
\$DerivationParameter	Coded term or Context Group for Concept Name of a derivation parameter
\$DerivationParameterUnits	Units of derivation parameter

FIGURE 4.2 – Les paramètres du concept "ROI measurements" du template "TID 1410" [Clunie 2012]

**Context ID RPH\_7601**  
**Measurement Object Types**  
Type: Extensible Version: 20090702

Coding Scheme Designator	Code Value	Code Meaning
99RPH	RP-100103	Line Segment
99RPH	RP-100104	Biorthogonal Line Segments
99RPH	RP-101016	Triorthogonal Line Segments
99RPH	RP-100105	Angle
DCM	111041	Outline
SRT	G-D705	Volume
99RPH	RP-100121	3D Geometric Object
99RPH	RP-100106	Text Annotation
99RPH	RP-100111	Measurements Derived From Other Measurements

FIGURE 4.3 – Les types de l'entité "measurement objects" (tableau "CID RPH7601 measurement object types" [Clunie 2012])



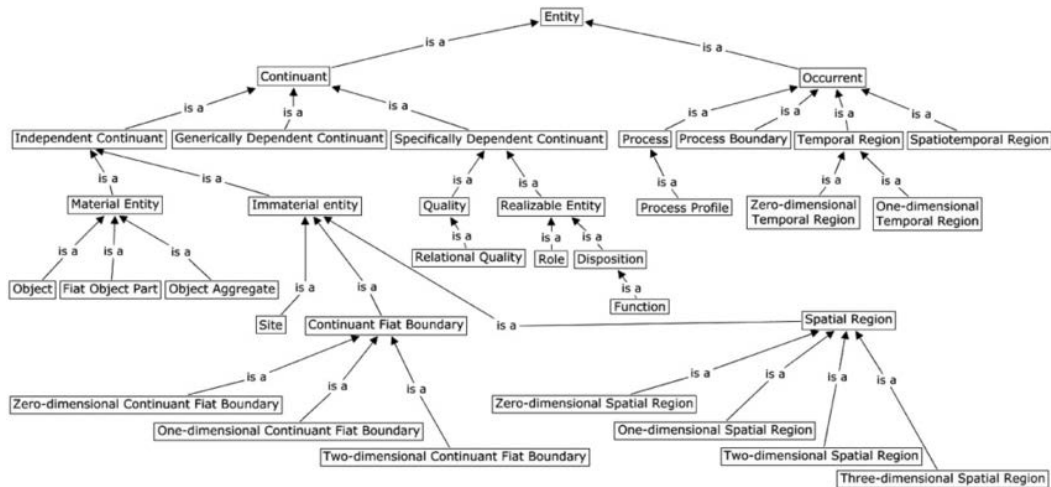


FIGURE 4.4 – La hiérarchie des classes de l'ontologie BFO [Smith 2012]

une structure anatomique ou un autre objet) de cette qualité (par exemple la taille), la mesure dépend de la façon dont elle a été concrétisée (par exemple, la procédure de calcul). Dans BFO, les entités de type "BFO:specifically dependent continuant" peuvent être des rôles, des qualités ou bien des dispositions et les entités "BFO:generically dependent continuant" font référence à des artefacts d'information (tels que des fichiers, des documents, des mesures etc.). La principale distinction entre une dépendance spécifique et une dépendance générique est que les "BFO:specifically dependent continuant" dépendent d'un "BFO:independent continuant" pendant toute la durée de son existence alors que les "BFO:generically dependent" continuant" dépendent d'une ou de plusieurs autres entités.

- Les classes et les instances : les classes représentent des concepts abstraits (par exemple, cerveau, lésion, etc.) et les instances représentent des objets particuliers (par exemple, le cerveau de ce patient, la lésion de David, etc.). C'est l'approche commune des ontologies en format OWL.

Ces différentes catégorisations sont représentées dans la Figure 4.4 qui donne une vue générale sur l'ensemble de la hiérarchie de l'ontologie BFO 2. La structure de l'ontologie BFO et d'autres détails sont fournis dans ce livre [Arp 2015]. La Figure 4.4 donne des exemples de type de classes et d'instances des trois catégories "BFO:independent continuant", "BFO:dependent continuant" et "BFO:occurrent".

#### 4.3.2.1 Version 1 de BFO

Les relations de la version 1 de BFO<sup>2</sup> sont définies dans l'ontologie RO (Relation Ontology). L'ontologie RO permet de représenter un ensemble de relations basiques qui sont très utilisées dans le domaine biomédical [Smith 2005a]. On peut citer à

2. <http://purl.obolibrary.org/obo/bfo.owl>

Niveaux de modélisation	Relations	Exemples de triplet
<universel, universel>	isA, partOf, ...	"hand part-of body"
<particulier, particulier>	instanceOf	"David's hand part-of David"
<particulier, universel>	partOf, ...	"David's hand instance-of hand"

TABLE 4.1 – Les trois niveaux de relations de l'ontologie RO de la fonderie OBO

titre d'exemple les relations "R0:participatesIn", "R0:partOf", "R0:locatedIn", etc. Ces différentes relations sont catégorisées par RO selon l'approche réaliste de modélisation [Smith 2006a], d'une manière plus simplifiée l'ontologie RO distingue entre deux niveaux de modélisation : le niveau d'abstraction et le niveau de spécialisation. Ainsi, elle définit trois types de relations binaires (voir le tableau 4.1). La plupart des relations définies dans RO ont une relation inverse correspondante [Smith 2007]. Par contre, le domaine et la portée des relations ne sont pas définis dans RO.

#### 4.3.2.2 Version 2 de BFO

La version 2 de BFO<sup>3</sup> est apparue en 2012, son objectif est d'intégrer les classes de BFO1 et les relations de RO dans une seule ontologie et d'enrichir les relations ainsi que les classes en spécifiant les contraintes sur leurs extrémités. Actuellement, la version 2 de BFO contient 35 classes, 36 propriétés d'objets, 0 propriétés de données et 2 propriétés d'annotation. La version 2 de BFO a remplacé les simples relations de RO par des relations temporelles, par exemple la relation "R0:partOf" est remplacée par les deux relations "BFO:partOfContinuantAtAllTimes" et "BFO:partOfContinuantAtSomeTime". Plusieurs problèmes ont été soulevés dans l'utilisation des relations temporelles de BFO2, ces problèmes sont discutés dans ce papier [Mungall 2013].

D'autre part, la version 2 de BFO a introduit le concept de "BFO:process profile" pour annoter des processus de mesurage qui estiment des qualités dans un intervalle de temps. A titre d'exemple, le processus de mesurage de la fréquence cardiaque (pulse rate) nécessite la qualification du processus de battement du cœur dans un intervalle de temps donné (voir Figure 4.5). La classe "BFO: process profile" subsume trois classes qui sont : la classe "BFO:quality process profile", la classe "BFO:rate process profile" et la classe "BFO:beat process profile".

Les deux versions de l'ontologie BFO sont incompatibles même si la hiérarchie des classes est la même. Certaines ontologies de la librairie OBO se sont alignées sur l'ontologie BFO2 en n'utilisant que ses classes et pas ses relations temporelles (dont la définition n'est pas encore stable dans le modèle) ; on peut citer à titre d'exemple les ontologies OGMS, OBI et IAO. On note que BFO2 n'a pas préservé les identifiants attribués par BFO1, les classes de BFO2 utilisent l'espace de noms d'OBO.

3. <http://purl.obolibrary.org/obo/bfo.owl>

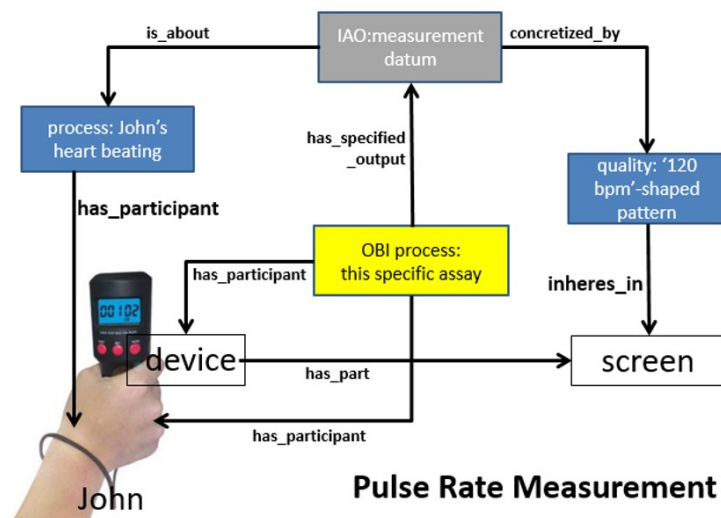


FIGURE 4.5 – Le problème de modélisation des processus de mesure de qualités temporelles ("John's heart beating").(extraite de la présentation "BFO tutorial" de Barry Smith à la conférence icbo 2011)

Ainsi, la classe racine qui est nommée "BFO:Entity" dans BFO1 est référencée sous BFO2 avec le nom de "BFO\_0000001". Dans notre travail, nous avons utilisé la version complète de BFO2 pour développer notre ontologie IBO.

### 4.3.3 Réutilisation des ontologies en fonction des contextes

Nous avons spécialisé de nombreuses classes qui sont issues d'ontologies existantes afin de les adapter à notre domaine d'application qui est, rappelons-le, la représentation des trois axes sémantiques relatifs au concept de biomarqueur d'imagerie (voir Figure 4.6). En particulier, nous avons utilisé les ontologies suivantes :

- L'ontologie OBI (Ontology for Biomedical Investigations) [Bandrowski 2016] pour représenter les données liées au processus de mesure ;
- L'ontologie PATO (Phenotypic Quality Ontology) [Mungall 2007] pour décrire les qualités liées au concept de biomarqueur d'imagerie (la taille, la forme, la concentration, etc.) ;
- L'ontologie FMA (Foundational Model of Anatomy) ontology [Rosse 2003] pour la représentation des structures anatomiques ;
- L'ontologie UO (Unit Ontology)<sup>4</sup> pour représenter les unités de mesures des biomarqueurs de type scalaire ;
- L'ontologie OntoNeuroLOG (Data-set (ONL-DS) et Data-set Processing (ONL-DP)) pour représenter non seulement la classification des images médicales (anatomique, fonctionnelle, métabolique) mais aussi la classification des traitements en imagerie médicale (recalage, ré-échantillonnage, segmentation, estimation des paramètres quantitatifs, etc.) ;

4. <http://purl.bioontology.org/ontology/UO>

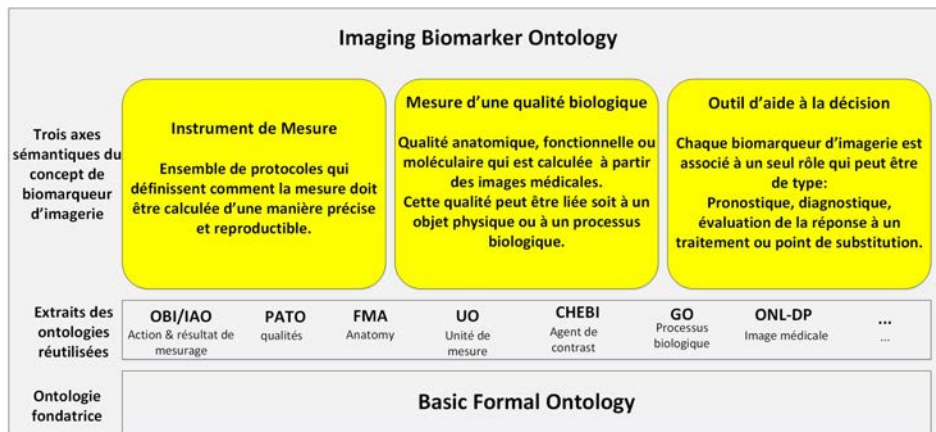


FIGURE 4.6 – La structure générale de l'ontologie IBO

- L'ontologie QIBO pour la modélisation de quelques concepts relatifs à la mesure ainsi que l'utilisation de la valeur du biomarqueur d'imagerie ;
- L'ontologie GO (Gene Ontology) [Consortium 2008] pour la description des processus biologiques ;
- L'ontologie ChEBI (Chemical Entities of Biological Interest Ontology) [Degtyarenko 2008] pour décrire les produits de contraste ;
- L'ontologie HDO (Human Disease Ontology) [Schriml 2011] pour représenter les pathologies (maladie neuronale, maladie cardiovasculaire, etc.) étudiées par les biomarqueurs d'imagerie.

Nous avons fait notre sélection d'ontologies en tenant comptes principalement des aspects suivants : (1) la disponibilité gratuite de l'ontologie sur le web, (2) la bonne définition des termes et des relations de l'ontologie afin d'être sûr de la réutilisation adéquate des parties extraites de l'ontologie, (3) la couverture du domaine visé pour créer un minimum d'ensemble de termes pour couvrir notre cas d'utilisation et (4) la stabilité de l'ontologie pour que les futures modifications n'affectent pas notre modèle (généralement se sont les ontologies les plus stables qui sont adoptées par les projets).

#### 4.3.3.1 Réutilisation de l'ontologie OBI

L'ontologie OBI (Ontology for Biomedical Investigations)<sup>5</sup> [Bandrowski 2016] définit environ 2360 termes et 40 relations pour représenter les processus expérimentaux dans le domaine biomédical : les protocoles expérimentaux, les matériels utilisés, les données générées (les mesures, les conclusions d'investigations, les valeurs de prédiction, etc.), les objectifs de l'étude, les analyses réalisées, etc. L'objectif d'OBI est de standardiser la définition des expérimentations pour pouvoir les comparer, les reproduire et ainsi améliorer leur intégration avec les autres données du web. OBI réutilise les classes de l'ontologie BFO et les relations de BFO et de RO.

5. <http://purl.obolibrary.org/obo/obi.owl>

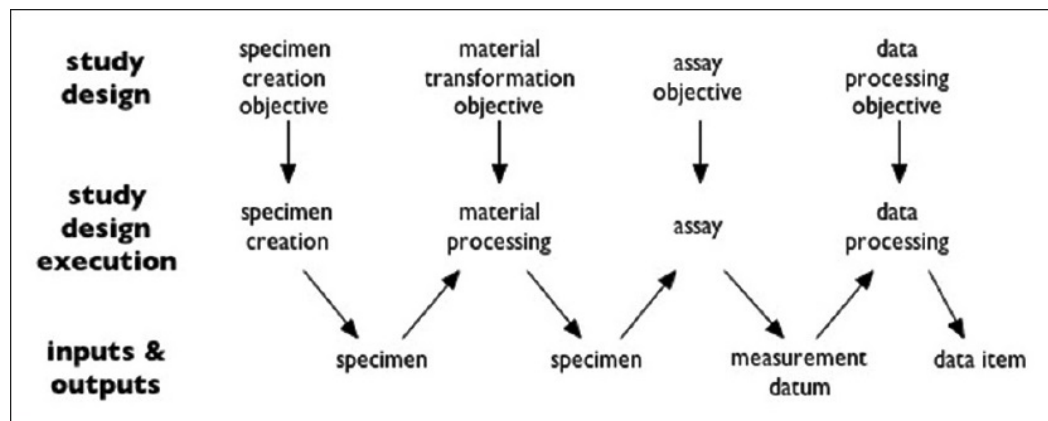


FIGURE 4.7 – La modélisation d’une investigation selon l’ontologie OBI [Smith 2015]

OBI étend l’ontologie IAO et fait partie des principales ontologies de la bibliothèque OBO.

Après une analyse du contenu de l’ontologie OBI, nous avons trouvé que les termes d’OBI répondent à certains de nos principaux besoins de modélisation et qu’ils sont suffisamment bien définis pour être réutilisés (présence d’annotations, de description formelle et d’exemples pour l’illustration de l’utilité des termes) : taxonomie liée aux mesures (valeur scalaire, valeur nominale, etc.), processus planifiés (acquisition de l’image, injection d’un objet matériel dans le corps humain), calculs statistiques, rôles (patient, participant à une investigation, groupe d’étude, etc.), critère d’éligibilité, et d’autres termes.

La figure 4.7 illustre la structure générale du processus d’investigation selon l’ontologie OBI. Les concepts définissant le processus d’investigation sont basés sur l’ontologie BFO : "BFO:process" pour la description des essais et des traitements (le niveau "study design execution" dans la figure 4.7), "BFO:material entity" pour modéliser les entrées et sorties des processus expérimentaux (le niveau "inputs/outputs" dans la figure 4.7) et "BFO:dependent continuant" pour la spécification des caractéristiques intrinsèques aux entités liées à la définition du processus d’investigation (par exemple, "study design" dans la figure 4.7). La figure 4.8 décrit le processus de mesure de la valeur du glucose chez un sujet biologique (d’autres exemples de cas d’application d’OBI sont décrits dans ce papier [Brinkman 2010]).

Actuellement, l’ontologie OBI est la ressource terminologique la plus adaptée à l’expression des investigations dans le domaine de la cancérologie. Elle a été réutilisée dans l’ontologie OBIB (Ontology for Biobanking) pour décrire d’une manière sémantique les méta-données associées aux bio-banques (par exemple, spécimen des tumeurs cancéreuses, données génomiques, etc.).

Pour notre travail, nous avons principalement réutilisé les entités suivantes d’OBI : "OBI:assay objective", "OBI:data transformation objective", "OBI:study design", "OBI:value specification", "OBI:genetic

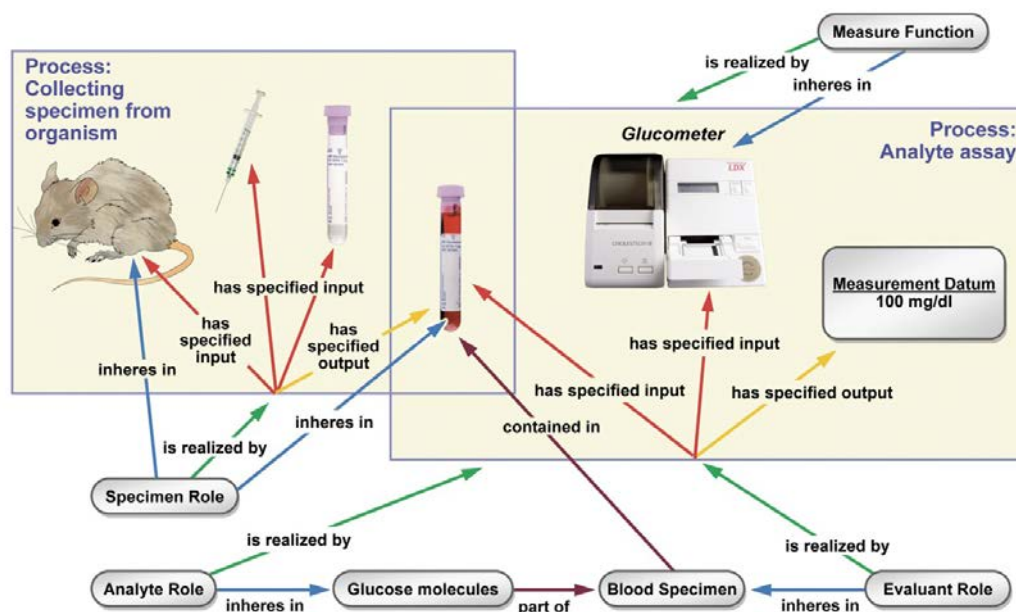


FIGURE 4.8 – Représentation du processus de mesure de la glycémie avec l'ontologie OBI, [Bandrowski 2016]

characteristics information", "OBI:dose", "OBI:study group role", etc. Et pour la représentation des "OBI:planned process", nous avons retenus les termes comme "OBI:assay", "OBI:data transformation", "OBI:investigation", "OBI:investigation", "OBI:material processing", etc.

#### 4.3.3.2 Réutilisation de l'ontologie IAO

L'ontologie IAO (Information Artifact Ontology)<sup>6</sup> [Ceusters 2012] modélise les ressources informationnelles d'une manière indépendante du domaine. Cette ontologie a été initialement créée pour répondre aux besoins exprimés par les développeurs d'OBI de fonder un vocabulaire contrôlé pour décrire les éléments informationnels liés aux expérimentations scientifiques. Ces données peuvent concerner par exemple les protocoles, les bases de données, les résultats de mesure, etc. C'est la classe centrale d'IAO appelée "IAO:information content entity" (ICE) qui modélise ces différents types de contenu informationnel d'une entité d'autres entités.

L'ontologie IAO définit 109 classes et 39 propriétés (16 propriétés d'objet, 4 propriétés de données, 19 propriétés d'annotation) et elle est fondée sur la version 2 de BFO. La classe "IAO:information content entity" comporte 102 sous-classes ("IAO:data item", "IAO:datum label", "IAO:figure", etc.). Pour représenter par exemple l'expression suivante "la largeur d'une tumeur est de 7cm", l'ontologie IAO propose d'utiliser la classe "IAO:length measurement datum" qui

6. <http://purl.obolibrary.org/obo/iao.owl>

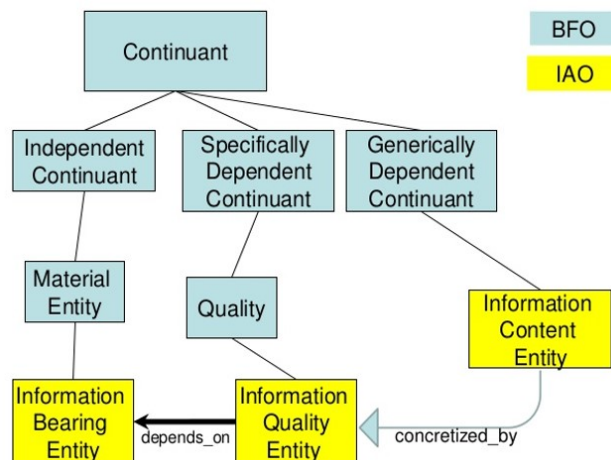


FIGURE 4.9 – L’alignement de trois classes principales de l’ontologie IAO à l’ontologie BFO

définit les mesures scalaires qui estiment la longueur, la classe "IAO:length" précise la qualité quantifiée et la classe "IAO:measurement unit length" regroupe les différentes unités de mesure qui peuvent être utilisées pour mesurer la longueur. La relation "IAO:hasMeasurementUnitLabel" lie les deux classes "UO:measurement unit length" et "IAO:length measurement datum", et la relation "OBI:hasSpecifiedValue" sert à spécifier la valeur de la mesure.

L’ontologie IAO a été réutilisée par plusieurs autres ontologies ; on peut citer à titre d’exemples : EFO, VO (Vaccine Ontology), OGMS, IDO (Infectious Disease Ontology), OCRE (Ontology of Clinical Research), etc (ces ontologies sont disponibles sur le NCBO BioPortal). Dans notre travail, nous avons utilisé l’intégralité de l’ontologie IAO (nous avons supprimé les classes obsolètes dans la version OWL de l’ontologie) qui est déjà réutilisée dans l’ontologie OBI.

#### 4.3.3.3 Réutilisation de l’ontologie UO

L’ontologie UO (Unit Ontology)<sup>7</sup> définit 381 classes pour la formalisation de la représentation des unités de mesure dans le domaine biomédical. L’ontologie UO a été développée pour associer aux observations quantitatives en biologie une unité de mesure [Gkoutos 2012]. La taxonomie des classes de l’ontologie UO couvre les préfixes et les unités de mesure.

- la classe "UO:prefix" représente les préfixes (milli, nano, etc.) associés aux unités de mesures.
- la classe "UO:unit" définit la hiérarchie des différentes unités de mesure associées aux qualités "PATO:quality" via la relation d’objet "PATO:isUnitOf". Par exemple, les sous classes de la classe "UO:length unit" sont

7. <http://purl.bioontology.org/ontology/UO>

"UO:centimeter", "UO:meter", "UO:millimeter", etc. et décrivent la qualité "PATO:1-D extent" qui est la sous-classe de l'entité "PATO:size".

Dans notre travail, nous avons retenu que les entités suivantes : "UO:length unit", "UO:radioactivity concentration", "UO:area unit", "UO:mass density unit", "UO:volume unit", "UO:volumetric flow rate unit", "UO:activity (of a radionuclide) unit", "UO:mass unit", "UO:acceleration unit" et nous avons défini ces entités sous la classe "IAO:measurement unit label".

#### 4.3.3.4 Réutilisation de l'ontologie FMA

L'ontologie FMA (Foundational Model of Anatomy) [Rosse 2003] est une ontologie de référence pour la modélisation des structures anatomiques du corps humain ainsi que des descripteurs des entités physiques (surfaces, lignes, objets, points, etc.) dans le domaine biomédical. Elle a pour vocation selon ses auteurs à être réutilisée en partie et adaptée à un domaine spécifique. FMA contient 104522 classes de représentation des entités anatomiques matérielles et immatérielles et 168 relations, ce qui fait de FMA l'une des plus "grandes" ressources ontologiques dans le domaine biomédical. L'ontologie est composée de quatre composantes inter-liées qui sont :

- La composante AT (Anatomy Taxonomy) organise les entités anatomiques physiques et non physiques en tenant compte de leur structure, c'est à dire des parties anatomiques qu'elles les composent (Figure 4.10). Ainsi, les entités anatomiques de niveau générique sont liées à leurs sous parties anatomiques (concepts spécifiques) via les relations "IsA". FMA distingue entre trois sous ensembles de parties anatomiques : (1) des macromolécules biologiques, des cellules et leurs parties, des portions de tissus, (2) des organes et leurs parties, ainsi que (3) des systèmes organiques et des parties du corps humain.
- La composante ASA (Anatomical Structural Abstraction) spécifie les relations spatiales et d'appartenance ("FMA:partOf" et son inverse "FMA:hasPart") pour établir les liens entre les différentes entités de la composante AT. La relation "FMA:partOf" et son inverse joue un rôle important dans la représentation des connaissances anatomiques et elle spécialise des relations génériques et spécifiques ("FMA:hasRegionalPart", "FMA:hasSystemicPart" et "FMA:constitutional"). Par exemple, "FMA:Neuron" "FMA:hasConstitutional" "FMA:plasma membrane", "FMA:Neuron" "FMA:hasRegionalPart" "FMA: cell body" et "FMA:hand" "FMA:hasConstitutional" "FMA:skin of hand".
- la composante ATA (Anatomical Transformation Abstraction) définit les transformations liées aux entités de la composante AT au cours des deux cycles pré-natale et post-natale.
- La composante MK (Meta-knowledge) spécifie les règles et les définitions des entités ainsi que des relations de FMA.

Nous avons réutilisé l'ontologie FMA pour représenter les structures anatomiques qui sont représentées dans les images médicales. Nous n'avons pas utiliser l'intégralité



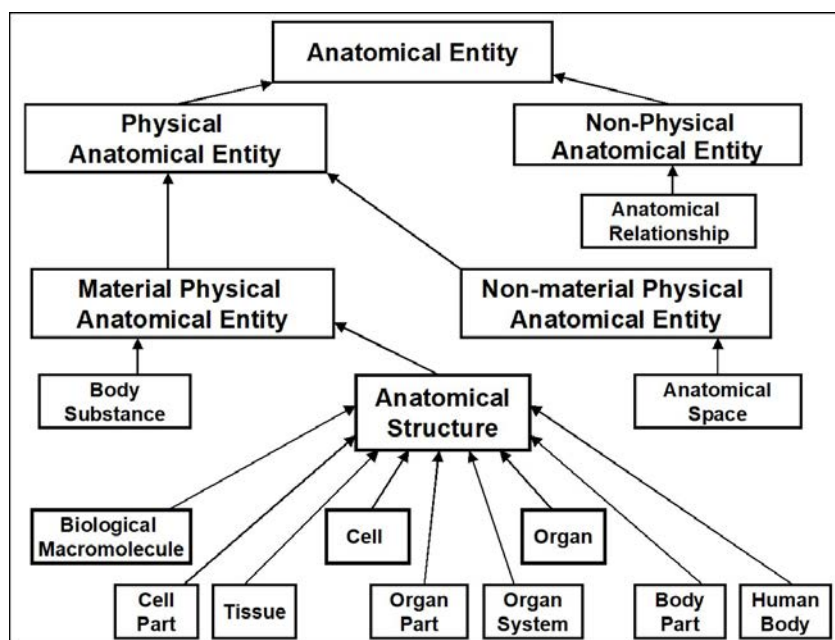


FIGURE 4.10 – Classification générale des entités anatomiques dans l'ontologie FMA, extraite de la présentation de Barry Smith sur BFO2

de FMA, nous nous sommes basés sur le tableau "Context ID RPH\_7006c anatomic sites for lesions" dans [Clunie 2007b] qui liste les entités anatomiques décrites dans les comptes rendus radiologiques dans le cadre des essais cliniques afin de repérer les principales entités qu'on doit inclure. Ainsi, IBO fait référence à "FMA:liver", "FMA:pancreas", "FMA:breast", "FMA:neck", "FMA:pelvis", "FMA:brain", etc.

#### 4.3.3.5 Réutilisation de l'ontologie PATO

L'ontologie PATO (Phenotypic Quality Ontology)<sup>8</sup> [Mungall 2007] est une ontologie qui est utilisée conjointement avec d'autres ontologies (d'anatomie (FMA), de biologie (GO), etc.) pour décrire les propriétés des entités physiques ou de processus. Dans PATO, une propriété est définie en anglais de la façon suivante "a dependent entity that inheres in a bearer by virtue of how the bearer is related to other entities". La classe "PATO:quality" regroupe les différentes catégories d'entités qui décrivent :

- Des propriétés liées à un objet physique ("PATO:physical object quality") : morphologiques (taille, longueur, forme, etc.), physiques (masse, couleur, etc.), etc. ;
- Des propriétés qualitatives ("PATO:qualitative") : "PATO:count", "PATO:magnitude", "PATO:intensity", etc. et d'autres propriétés plus complexes qui décrivent la déviation de la valeur de la qualité par rapport

8. <http://purl.obolibrary.org/obo/pato.owl>

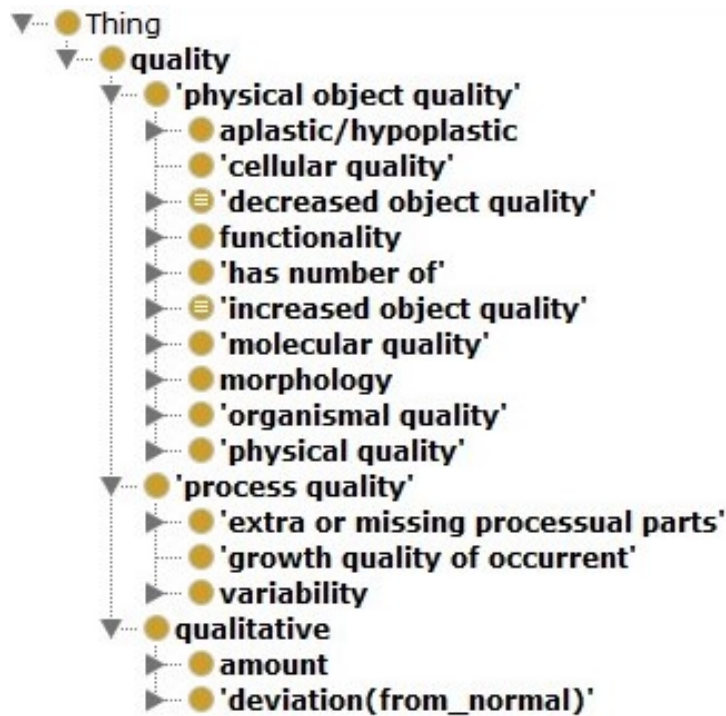


FIGURE 4.11 – Concepts extraits de l'ontologie PATO

à la normalité. Ces classes sont introduites sont nommées "PATO:decreased quality" et "PATO:increased quality".

- Des propriétés liées aux processus ("PATO:process quality") : par exemple pour la qualité "PATO:color" ces différentes qualités lui sont aussi attribuées "PATO:variability of color", "PATO:increased variability of color", "PATO:decreased variability of color".

Comme nous l'avons expliqué dans le chapitre introductif de la thèse, les biomarqueurs d'imagerie sont utilisés pour évaluer des propriétés anatomiques, morphologiques ou physiologiques. Pour représenter ces différentes qualités (quantifiables et non quantifiables) mesurées à l'aide des biomarqueurs d'imagerie nous avons choisi d'employer quelques termes de chaque une des trois catégories citées ci-dessus. Figure 4.11 montre les principaux concepts retenus. Nous précisons que nous avons réutilisé les termes de PATO pour la modélisation des propriétés temporelles comme la classe "BFO:quality process profile" n'est pas encore intégrée dans l'hierarchie des classes de BFO2. Nous avons aligné certaines relations de PATO (par exemple, "PATO:towards" correspond à "BFO:isAbout"), d'autres sont maintenues telle que la relation de PATO issues de BFO/RO "PATO:differentInMagnitudeRelativeTo" et ses descendants (pas de correspondant dans BFO).

#### 4.3.3.6 Réutilisation de l'ontologie OGMS

L'ontologie OGMS<sup>9</sup> définit les principales entités (environ 100 classes) pour décrire les symptômes pré-cliniques, les signes cliniques, les processus pathologiques ainsi que les actes cliniques (diagnostic, pronostic, etc.). L'objectif de l'ontologie OGMS est notamment de créer une terminologie qui permet de lever l'ambiguïté autour de la définition du terme "disease". Ainsi OGMS distingue dans sa classification deux aspects importants qui sont : (1) les observations cliniques mesurées effectuées par les cliniciens et (2) les processus biologiques qui existent réellement dans l'organisme du patient. L'apport de cette ontologie dans le domaine biomédical consiste en la formalisation des entités cliniques intervenant dans l'évaluation clinique de l'état du patient.

L'ontologie OGMS est basée sur l'ontologie BFO 2 et elle respecte les principes de la fonderie OBO. OGMS définit le concept "OGMS:disease" comme une entité de type "BFO:disposition" qui est causée par des troubles physiques et qui évolue dans un processus pathologique appelé "OGMS:disease course". Ainsi, la maladie comme toute entité de type disposition peut être réalisée dans un processus. Par conséquent, le terme "OGMS:disease course" est défini de la façon suivante : "disease course is the totality of all processes through which a given disease instance is realized".

OGMS répartit ses classes en quatre grandes catégories [Scheuermann 2009] (voir Figure 4.12) :

- "OGMS:symptom" : est une caractéristique corporelle qui est subjectivement observée par le patient et qui la considère comme une preuve d'une maladie. Par exemple, la mesure de la température corporelle sans l'emploi d'un thermomètre.
- "OGMS:sign" : est une caractéristique corporelle qui est objectivement mesurée à travers un examen médical et qui a une signification clinique. Par exemple, la mesure de la taille d'une lésion.
- "OGMS:disorder" : cette entité fait référence à l'ensemble maximal de plusieurs composantes pathologiques. Par exemple une tumeur est un "OGMS:disorder" a elle est composée d'un ensemble maximal de cellules anormales. Contrairement à la maladie, elle est de type matériel et non pas une disposition.
- "OGMS:clinical finding" : est la représentation d'une évaluation clinique de plusieurs examens médicaux physiques ou d'imagerie. Par exemple, tumeur visualisée à partir des examens d'imageries.
- "OGMS:diagnosis" et "OGMS:prognosis" : sont définis comme un "BFO:continuant" qui représentent des résultats concluants qui sont issus de la réalisation d'un acte clinique.

Dans notre travail, nous n'avons réutilisé que quelques entités de l'ontologie OGMS qui sont de type "BFO:continuant" et nous avons aligné les relations de RO à l'ontologie BFO2 (voir Figure 4.13). Nous avons réutilisé les classes de l'ontologie

---

9. <http://purl.obolibrary.org/obo/ogms.owl>

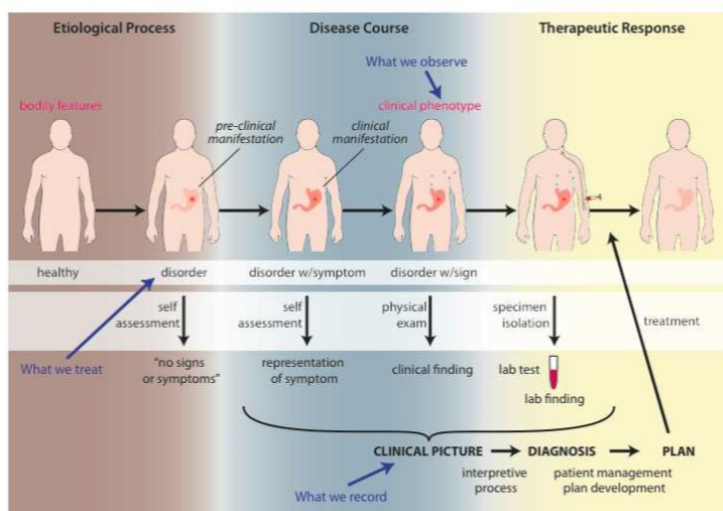


FIGURE 4.12 – Schématisation de l'évolution temporelle de la maladie au sein d'un organisme de l'étape pré-clinique à l'étape de traitement. Toutes les définitions des termes sont données dans ce papier [Scheuermann 2009]

OBI pour la définition des processus de diagnostic et de pronostic vu qu'ils sont mieux formellement définies dans OBI. Nous n'avons pas inclus des termes comme, à titre d'exemple, la classe "OGMS:clinical picture" qui est définie de façon peu précise : (1) elle ne fait pas référence à l'aspect de mesure d'une qualité physiologique (anatomique, moléculaire, etc.) et (2) elle n'a pas été spécifiée pour couvrir les images médicales issues de différents appareils d'imagerie médicale (IRM, PET, etc.).

#### 4.3.3.7 Réutilisation de l'ontologie QIBO

L'ontologie QIBO est décrite en détail dans le Chapitre de l'état de l'art. Dans le développement du modèle IBO, nous avons retenu quelques classes de QIBO : la classe "QIBO:imaging agent" et ses sous-classes, la classe "QIBO:imaging subject" et ses sous-classes, la classe "QIBO:disease diagnostic", la classe "QIBO:disease staging", la classe "QIBO:treatment assessment", la classe "QIBO:screening", la classe "QIBO:prognosis" et la classe "QIBO:predicting response to therapy". Nous avons alignée ces différentes classes à des termes provenant de l'ontologie OBI ou bien directement sous les termes de l'ontologie BFO (voir Tableau 4.2).

Nous avons pris en considération les principaux axes sémantiques de QIBO pour identifier les classes fondamentales du domaine des biomarqueurs. Ainsi, nous avons essayé de chercher des classes des ontologies d'OBO qui peuvent être réutilisées pour la conceptualisation de ce que QIBO cherche à représenter. Le Tableau 4.3 illustre le travail de couverture sémantique que nous avons réalisé.

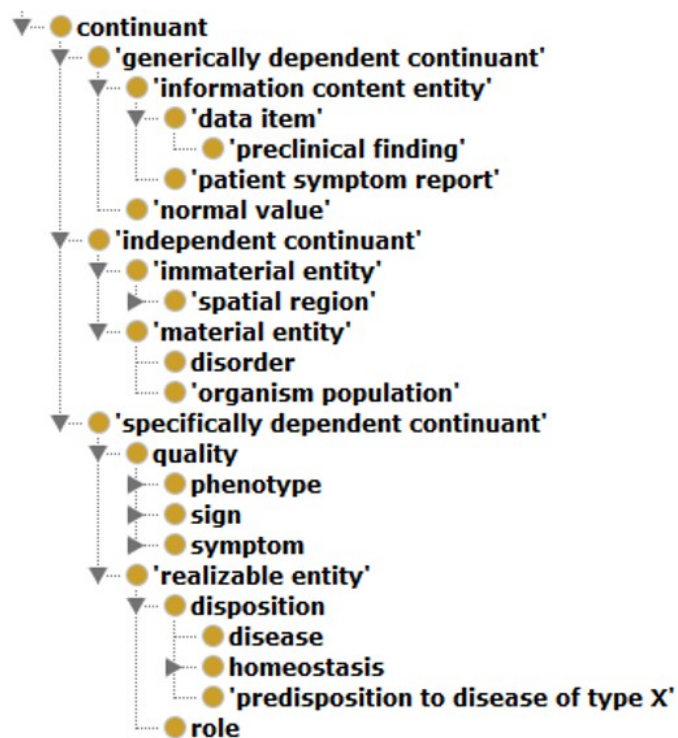


FIGURE 4.13 – Les entités extraites de l'ontologie OGMS

Classe QIBO	Classes de la fonderie OBO
QIBO:imaging agent	OBI:material to be added
QIBO:imaging subject	BF0:material entity
QIBO:disease diagnostic	OBI:assay
QIBO:disease staging	OBI:assay
QIBO:treatment assessment	OBI:planned process
QIBO:Predicting response to therapy	OBI:prediction
QIBO:screening	OBI:performing a diagnosis
QIBO:prognosis	BF0:planned process

TABLE 4.2 – Alignement des termes de QIBO aux classes d'OBO

Classes QIBO	Correspondance	Représentation dans IBO
QIBO:biological target	$\subseteq$	OBI:target of material addition
QIBO:biological intervention	=	OBI:material processing
QIBO:biomarker use	$\subseteq$	OBI:planned process
QIBO:imaging agent	$\subseteq$	OBI:material to be added
QIBO:imaging subject	$\subseteq$	IBO:imaging subject role
QIBO:imaging technique	=	ONL-DP:dataset
QIBO:indicated biology	$\subseteq$	GO:biological process, OGMS:disease
QIBO:post-processing algorithm	$\subseteq$	OBI:data transformation
QIBO:quantitative imaging biomarker	$\subseteq$	OBI:measurement datum

TABLE 4.3 – Correspondance entre les termes génériques de QIBO et les termes d'IBO.

#### 4.3.3.8 Réutilisation de l'ontologie OntoNeurolog

Nous avons introduit l'ontologie OntoNeurolog dans le Chapitre de l'état de l'art du travail de la thèse. Nous avons aligné les termes des deux ontologies ONL-DS et ONL-DP à l'ontologie BFO de la façon suivante : les classes "ONL-DS:dataset", "ONL-DS:dataset metadata", "ONL-DS:dataset set of values" sont définies comme des sous-classes de "IAO:data set". Les processus de traitement "ONL-DP:dataset processing" sont définis comme des "OBI:data transformation". Pour mieux couvrir le domaine, nous avons créé une classe "IBO:region of interest" pour se référer à une partie du "ONL-DS:dataset".

D'autre part, nous avons remplacé les relations spécifiques à l'ontologie DOLCE [Masolo 2003] par des relations de l'ontologie BFO. Nous citons ici les principales correspondances que nous avons réalisées :

- la relation "OBI:hasSpecifiedInput" correspond à la relation "DOLCE:hasForDataAt";
- la relation "DOLCE:hasForQuality" a été remplacée par la relation "BFO:hasQualityAtAllTimes";
- la relation "DOLCE:hasForPartDuring" correspond à la relation "BFO:partOf";

#### 4.3.3.9 Réutilisation de l'ontologie GO

L'ontologie GO (Gene Ontology)<sup>10</sup> [Consortium 2008] est le résultat d'une initiative collaborative qui vise à la formalisation des terminologies de représentation du domaine biologique. L'ontologie GO contient 48977 termes et elle est divisée en trois grandes parties :

- "GO:cellular component" : décrit la localisation (partie d'une cellule ou de son environnement extracellulaire) dans laquelle se trouve le produit de gène. Par exemple, une cellule, une membrane, un axone neuronal, etc.
- "GO:molecular function" : décrit les activités à l'échelle moléculaire. Par exemple, une activité enzymatique, une activité du transporteur de neurotransmetteur, etc.
- "GO:biological process" : décrit les événements moléculaires qui réalisent des fonctions moléculaires "GO:molecular function" associées à des "GO:cellular component". Par exemple, la mort cellulaire, l'anaphase, la mitose, etc.

Dans notre travail, nous avons sélectionné des classes de la catégorie "GO:biological process" comme "GO:cell growth", "GO:cell division", "GO:cell cycle process". La sélection des termes à retenir n'était pas évidente pour nous vu que les processus biologiques mesurées par les biomarqueurs d'imagerie sont de nature très diversifiée. Nous soulignons donc le fait que la liste des classes extraites de l'ontologie GO n'est pas exhaustive et doit être étendue par les futurs utilisateurs d'IBO.

#### 4.3.3.10 Réutilisation de l'ontologie HDO

Nous avons réutilisé l'ontologie HDO (Human Disease Ontology) qui représente une classification étendue des maladies humaines par étiologie (voir Figure 4.14). L'ontologie DHO a adopté la définition du terme "disease" de l'ontologie OGMS : "a disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism". DHO inclut des termes de ressources ontologiques telles que la classification ICD et le lexique SNOMED CT pour couvrir le maximum de maladie de l'être humain qui sont employées dans le domaine biomédical. Nous avons exclu les termes descendants des classes "DOID:physical disorder" et "DOID:syndrome" (voir Figure 4.14).

#### 4.3.4 Construction des modules ontologiques

Dans la plupart des cas, seuls des sous-ensembles des ontologies citées dans la sous-section précédente sont nécessaires. Nous avons utilisé l'outil Ontofox<sup>11</sup> qui implémente la méthode MIREOT (Minimum Information to Reference an External Ontology Term) [Xiang 2010] pour l'extraction des parties pertinentes (classes, relations et annotations associées) des ontologies OBO suivantes : OBI, PATO, IAO,

10. <http://www.geneontology.org/>

11. <http://ontofox.hegroup.org/>

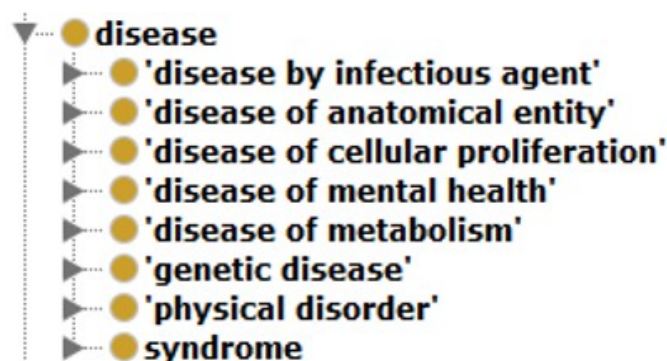


FIGURE 4.14 – L'hierarchie générale de l'ontologie DHO

UO, FMA, CHEBI et OGMS. Pour l'extraction d'un terme provenant d'une ontologie externe avec cette méthodologie, il faut préciser trois URIs dans le fichier d'entrée qui correspondent à :

- L'URI de l'ontologie source qui contient le terme qu'on veut importer ;
- L'URI du terme qu'on veut importer ;
- L'URI de la classe parente dans l'ontologie qui va faire référence au terme importé.

Des options sont disponibles pour permettre de spécifier des informations complémentaires à savoir :

- Classes : les descendants du terme à importer ("include all children") ;
- Axiomes : la définition formelle du terme ("includeAllAxioms"). L'option "includeAllAxiomsRecursively" permet d'importer les termes qui apparaissent dans la définition du terme importé ;
- Annotations : les définitions, les classes équivalentes, les labels, etc. ("includeAllAnnotationProperties" permet d'inclure tous les types d'annotations)

Les ontologies résultantes sont enregistrées dans des fichiers .owl et décrites en RDF/XML. On note que nous avons utilisé l'outil Protégé pour la génération des modules relatifs à l'ontologie ONL-DP et l'ontologie QIBO comme elles ne font pas encore partie de la fonderie OBO.

## 4.4 Résultat

### 4.4.1 Développement de l'ontologie IBO

L'ontologie IBO articule les trois aspects fondamentaux liés au concept de biomarqueur d'imagerie à savoir, la qualité biologique mesurée, l'outil de mesure et l'outil d'aide à la prise de décision. Les deux figures 4.15 et 4.16 montrent, respectivement, comment les entités de type "BFO:occurrent" et "BFO:continuant" qui composent les trois axes sémantiques ont été intégrées dans le cadre de l'ontologie BFO2.



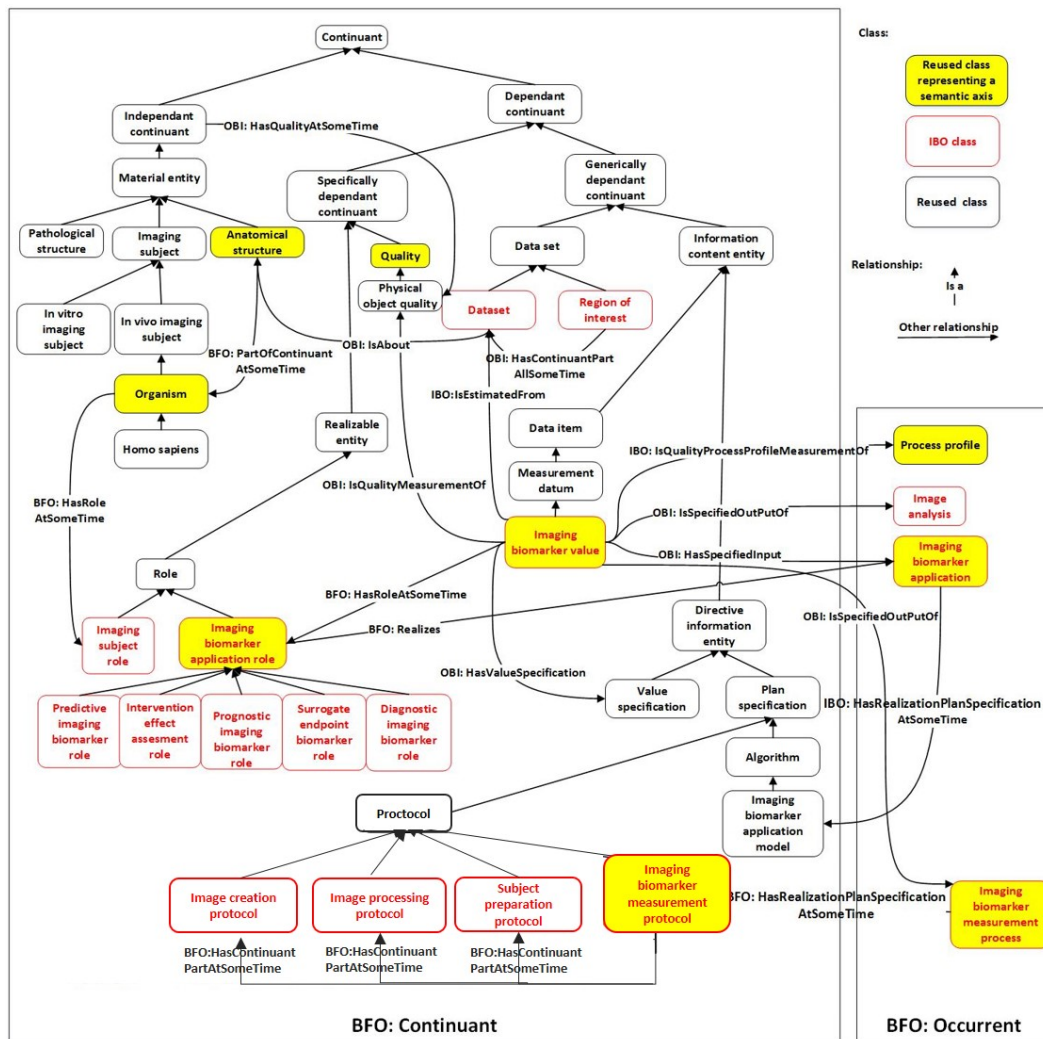


FIGURE 4.15 – Les principales composantes de type "BFO:continuant" de l'ontologie IBO. Les classes sont représentées avec des cadres et les relations entre les différentes classes sont dénotées par des flèches.

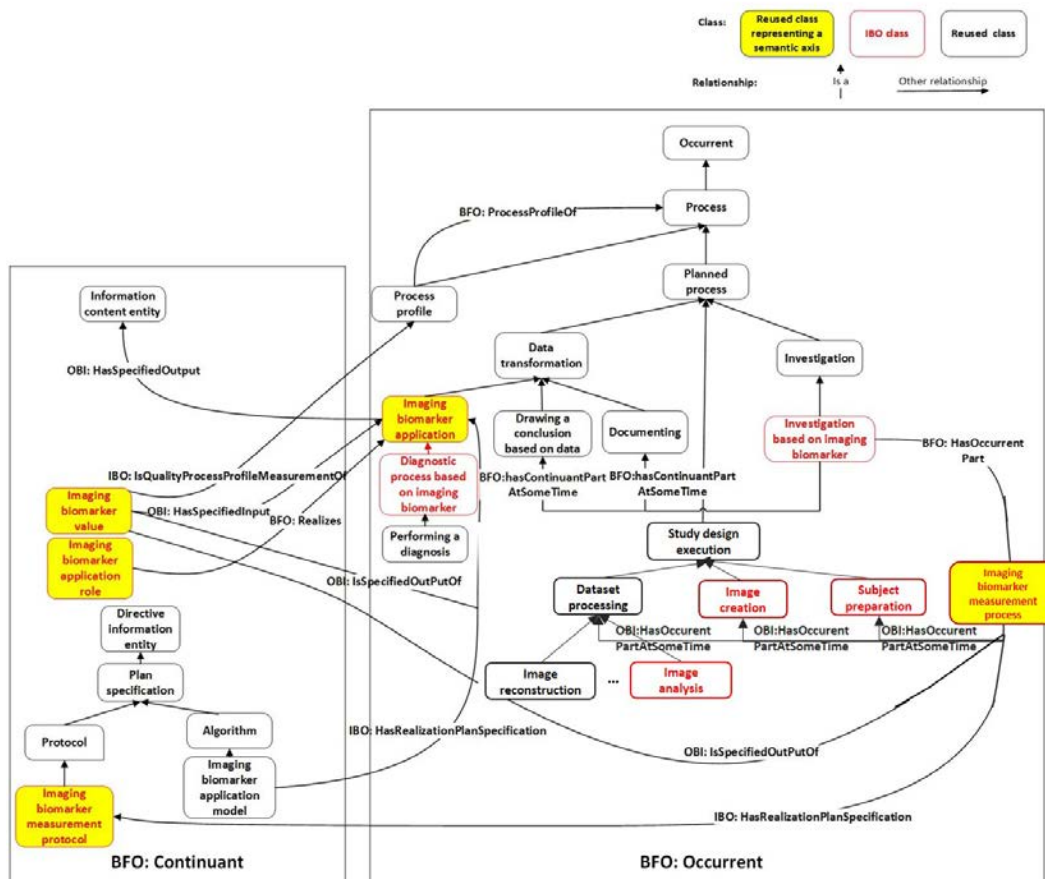


FIGURE 4.16 – Les principales composantes de type "BFO: occurrent" de l'ontologie IBO. Les classes sont représentées avec des cadres et les relations entre les différentes classes sont dénotées par des flèches.

#### 4.4.1.1 Aspect instrument de mesure

La notion de biomarqueur d'imagerie fait intervenir la classe "IBO:imaging biomarker measurement protocol" qui spécifie comment la valeur du biomarqueur d'imagerie doit être calculée d'une manière précise et reproductible. Le processus de calcul de la valeur de biomarqueur d'imagerie qui concrétise le plan qui a été spécifié par le protocole "IBO:imaging biomarker measurement protocol" est modélisé comme un "IBO:imaging biomarker measurement process". Cette définition de la classe "IBO:imaging biomarker measurement process" se traduit en langage OWL de la façon suivante : "IBO:imaging biomarker measurement process"  $\equiv$  def. "OBI:study design execution" and "BFO:realizes" some ("BFO:concretizes" some "IBO:imaging biomarker measurement protocol"). Un processus de type "IBO:imaging biomarker measurement process" "OBI:hasSpecifiedInput" some ("IAO:data item" or "OBI:organism").

Le processus de calcul de la valeur de biomarqueur d'imagerie est composé de différentes étapes, de trois types différents : "IBO:subject preparation", "OBI:image creation" qui modélise le processus d'acquisition de l'image médicale, les appareils d'imagerie utilisés et les participants au processus d'acquisition de l'image, etc., "ONL-DP:dataset processing" qui modélise les processus de traitement de l'image médicale comme par exemple IBO:image reconstruction, "ONL-DP:registration", "ONL-DP:restoration", "ONL-DP:segmentation", "IBO:image analysis". la classe "IBO:image analysis" modélise le traitement délivrant le résultat final du processus de mesure qui est la valeur de biomarqueur d'imagerie : "ONL-DP:quantitative image analysis" and "IBO:qualitative image analysis".

La modélisation de toutes ces différentes étapes citées ci-dessus fait intervenir plusieurs entités comme "OBI:device", "QIBO:imaging agent" et "FMA:anatomical structure", et les rôles joués par les différents participants aux processus de mesurage : "IBO:imaging subject role", "OBI:patient role", "OBI:material to be added role", etc.

#### 4.4.1.2 Aspect mesure d'une qualité biologique

La valeur de biomarqueur d'imagerie "IBO:imaging biomarker value" est décrite dans l'ontologie IBO de la façon suivante : "IBO:imaging biomarker value"  $\equiv$  def. "OBI:measurement datum" and "OBI:isSpecifiedOutputOf" some "IBO:imaging biomarker measurement process" and "OBI:isSpecifiedOutputOf" some "IBO:image analysis" and "OBI:hasRole" some "IBO:imaging biomarker application role". La valeur de biomarqueur d'imagerie subsume deux types de mesures : numériques et nominales. Cette spécification de la classification des mesures par leur nature est définie à travers l'entité "OBI:value specification". Dans la catégorie des biomarqueurs qualitatifs on trouve les valeurs nominales non ordonnées (par exemple, présence de kyste ("oui", "non"), aspect hémorragique ("oui", "non"), etc.) et les valeurs

nominales et ordonnées (par exemple, l'épaisseur de la marge ("aucune", "fine", "solide"), la définition du contour ("bien défini", "mal défini"), etc.). Dans la catégorie des biomarqueurs quantitatifs, on trouve les valeurs continues : les ratios, les valeurs avec unité de mesure, etc. et les valeurs discrètes : énumération. Les mesures des biomarqueurs quantitatifs sont le résultat de calcul du processus "ONL-DP:quantitative parameter estimation" qui fait partie d'une "IBO:quantitative image analysis" et les mesures des biomarqueurs d'imagerie qualitatifs sont le résultat de calcul du processus "IBO:qualitative parameter estimation" qui fait partie "IBO:qualitative image analysis".

Pour exprimer qu'un biomarqueur d'imagerie est une mesure d'une qualité biologique chez un patient, nous avons, d'abord lié la classe "IBO:imaging biomarker value" à la qualité de l'objet physique, elle même liée à cet objet physique (tumeur entière, partie de la tumeur, le contour de la tumeur, etc.) ou bien du processus métabolique qu'elle mesure (angiogenèse, mort cellulaire, etc.). Ensuite, nous avons distingué deux types de qualités : des qualités qui sont liées à des "BFO:continuant" (taille d'une tumeur, le volume d'une tumeur, présence ou non d'un kyste, etc.) et d'autres qui décrivent des "BFO:occurrent" (par exemple la restriction de la diffusion des molécules d'eau dans le tissu tumoral). Enfin, nous avons employé les deux relations "IBO:isProcessProfileMeasurementOf", "OBI:isQualityMeasurementOf", pour lier la valeur d'un biomarqueur d'imagerie à la qualité ou au processus estimé. Les qualités liées aux objets physiques sont modélisées comme des "PATO:quality" et celles qui sont liées aux processus sont modélisées comme des "BFO:process Profile". Nous avons spécialisé la classe "BFO:process Profile" en définissant la sous-classe "IBO:quality process profile" qui est introduite dans la documentation de la version 2 de BFO mais pas encore intégrée dans la version officielle. Cette classe va permettre de décrire la variation (augmentation ou diminution) de la qualité mesurée.

#### 4.4.1.3 Aspect outil d'aide à la décision

Selon cet axe sémantique, le biomarqueur d'imagerie est considéré comme un outil d'aide dans la prise de décision, par exemple établir la présence ou non d'une pathologie, la prédiction ou l'évaluation de la réponse à une thérapie. Nous avons défini cinq rôles que la classe "IBO:imaging biomarker value" peut jouer dans un processus décisionnel avec la classe "BFO:role". Tous les processus décisionnels sont de type "IBO:imaging biomarker application" et ils impliquent chacun un "IBO:imaging biomarker value" qui porte un "IBO:imaging biomarker role". La description du processus décisionnel est faite via la classe "imaging biomarker application" qui est définie de la façon suivante : "IBO:imaging biomarker application"  $\equiv$  def. ("BFO:realizes" some ("BFO:concretizes" some "IBO:imaging biomarker application model")) and ("OBI:hasSpecifiedInput" some ("IBO:imaging biomarker value" and ("BFO:hasRole" some "IBO:imaging biomarker application role"))) and ("OBI:hasSpecifiedOutput" some "IAO:information content entity")

## 4.4.2 Application de l'ontologie IBO pour la représentation de quelques biomarqueurs d'imagerie

### 4.4.2.1 Illustration dans le domaine des gliomes

Dans cette section, nous montrons comment l'ontologie IBO peut être utilisée pour représenter les observations d'imagerie de la collection de données TCGA (The Cancer Genome Atlas) GBM (glioblastoma multiforme) [TCGA-Research-Group 2014]. Ici, nous considérons l'étude rétrospective décrite dans ce papier [Jain 2014] qui se focalise sur la combinaison des biomarqueurs d'imagerie morphologiques et fonctionnels de la région de la tumeur GBM qui ne prend pas le contraste en anglais non-enhancing region ou NER. Le résultat des évaluations radiologiques est enregistré dans une feuille de calcul qui est disponible sur ce lien<sup>12</sup>. Cette étude émet l'hypothèse que les caractéristiques morphologiques de la région NER seules ne sont pas suffisantes pour prédire la survie du patient et que les paramètres de perfusion, à savoir la mesure du volume sanguin régional cérébral relatif (relative cerebral blood volume ou rCBV) de la région NER fournit une information de prédiction plus exacte.

**Exemple 1 : paramètre de perfusion rCBV** Le processus de quantification de la valeur moyenne de la mesure  $rCBV_{NER}$  à partir des images de perfusion cérébrale comprend trois étapes principales, à savoir : la préparation du sujet, l'acquisition et le traitement d'images. Au cours de la préparation du sujet, un agent de contraste IRM est utilisé pour visualiser la partie qui prend le contraste (en anglais contrast enhancing region ou CER). Par conséquent, un ensemble d'images de type "T2 star weighted perfusion images" a été généré à partir de l'étape d'acquisition de l'image. Les images acquises sont traitées pour corriger les fuites d'agents de contraste entre l'espace intravasculaire à l'espace extracellulaire. La valeur  $rCBV_{CER}$  est estimée à partir de la carte paramétrique rCBV. La figure 3 illustre la description sémantique du processus de mesure en utilisant le modèle IBO et le Tableau 4.4 présente certains détails sur les classes utilisées.

**Exemple 2 : critères VASARI** La terminologie VASARI (Visually Accessible Rembrandt Images) [TCGA-Research-Group 2014] est un vocabulaire d'annotation des gliomes cérébraux de haut grade en particulier le glioblastome multiforme (GBM) dans les images IRM. Son objectif principal consiste à normaliser la description des tumeurs cérébrales et à faciliter leur interprétation par les neuro-radiologues. La terminologie VASARI contient trente critères d'imagerie et elle a été développée par des experts du domaine qui ont considéré la majorité des évaluations possibles des tumeurs cérébrales en IRM. Chaque critère d'imagerie de la terminologie VASARI est référencé par un numéro ( $F_1$ ,  $F_2$ , etc.) et un ensemble de valeurs de scores possibles. Par exemple, le critère " $F_1$ :tumor location" de VASARI évalue la lo-

12. [https://wiki.cancerimagingarchive.net/download/attachments/18514300/JainPoisson\\_2014\\_Radiology\\_Dataset.xlsx?](https://wiki.cancerimagingarchive.net/download/attachments/18514300/JainPoisson_2014_Radiology_Dataset.xlsx?)

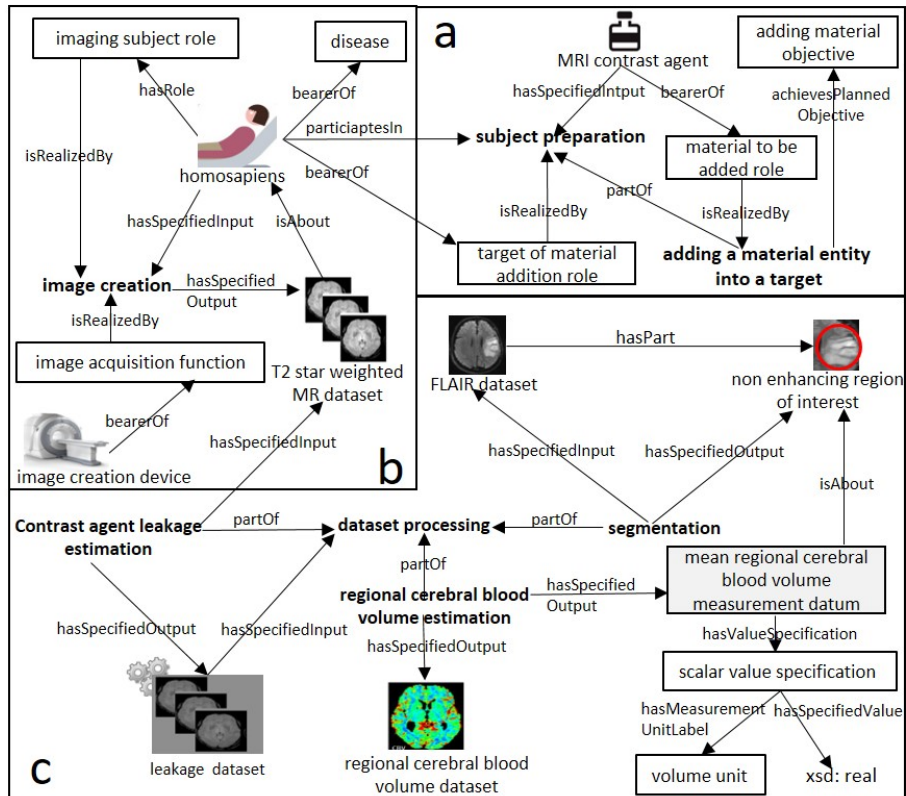


FIGURE 4.17 – Représentation du processus de mesure de la valeur rCBV avec l'ontologie IBO : le cadre "a" décrit les entités qui sont impliquées dans le processus de préparation du sujet, le cadre "b" correspond au processus de création d'image et le cadre "c" détaille l'ensemble de données du processus de traitement. Nous notons que les noms des processus sont mis en gras, les classes qui ne sont pas des processus sont contenues dans des cadres et les relations sont désignées par des arcs.

Classe	Super-classe
IBO:subject preparation	OBI:planned process
OGMS:disease	BFO:disposition
OBI:target to material addition role	BFO:role
OBI:material to be added role	BFO:role
OBI:MRI contrat agent	CHEBI:pharmaceutical
OBI:adding a material entity into a target	OBI:material combination
OBI:adding material objective	OBI:material combination objective
OBI:image creation	OBI:planned process
OBI:homo sapiens	OBI:organism
IBO:imaging subject role	BFO:role
ONL-DP:T2 star weighted MR dataset	ONL-DP:functional dataset, ONL-DP:MR dataset, ONL-DP:reconstructed-dataset
OBI:image acquisition function	OBI:measure function
OBI:image creation device	OBI:device
ONL-DP:dataset processing	OBI:data transformation
ONL-DP:segmentation	ONL-DP:dataset processing
IBO:non enhancing region of interest	ONL-DP:segmentation dataset
IBO:FLAIR dataset	ONL-DP:parameter quantification dataset
ONL-DP:regional cerebral blood volume estimation	ONL-DP:quantitative parameter estimation
ONL-DP:regional cerebral blood volume dataset	ONL-DP:parameter quantification dataset, ONL-DP:hemodynamic dataset
IBO:mean regional cerebral blood volume measurement datum	OBI:average value, IBO:imaging biomarker volume measurement datum
OBI:scalar value specification	OBI:numeric value specification
UO:volume unit	IAO:measurement unit label
IBO:qualitative parameter estimation	IBO:parameter estimation
IBO:non enhancing margin definition	IBO:categorical VASARI criterion
IBO:non enhancing margin definition value specification	IBO:categorical value specification
IBO:non enhancing margin definition label option	IBO:categorical label

TABLE 4.4 – Les classes utilisées dans les cas d'utilisation relatifs au domaine des gliomes

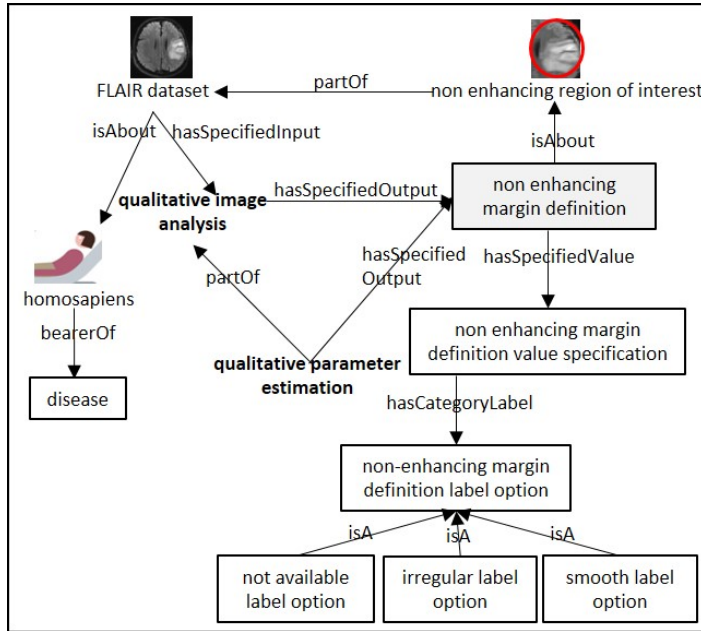


FIGURE 4.18 – La représentation du critère VASARI F<sub>29</sub> avec l’ontologie IBO. Nous notons que les processus sont représentés en gras, les classes qui ne sont pas des processus sont contenues dans des boîtes et les relations sont désignées par des arcs.

calisation de l’épicentre géographique et il définit six valeurs d’étiquette possibles = frontal, temporal, insulaire, pariétal, occipital, tronc cérébral, cervelet ; le critère "F<sub>29</sub>&F<sub>30</sub>:lesion size" de VASARI mesure le plus grand axe perpendiculaire à la plus petite dimension de l’anomalie qui est mesurée sur une seule image axiale (se référer au Tableau 4.4 pour voir les définitions de F<sub>1</sub> et F<sub>29</sub>&F<sub>30</sub> et d’autres exemples de critères VASARI).

Dans ce paragraphe nous montrons comment l’ontologie IBO permet de modéliser deux exemples de critères VASARI F<sub>13</sub> (voir Tableau 4.4 pour avoir quelques détails sur les classes utilisées) : F<sub>13</sub> décrit l’aspect de la marge de la région NER ("n/a", "smooth" ou "irregular") (voir Figure 4.18).

## 4.5 Discussion

Le résultat de notre travail montre qu’il est possible de représenter le concept de biomarqueur d’imagerie avec ses trois axes sémantiques en intégrant et spécialisant des classes provenant d’ontologies existantes. Nous avons donc créé une ontologie générique qui a comme objectif principal de définir d’une manière précise ce concept de biomarqueur d’imagerie et de lever l’ambiguïté sur ce terme afin d’assurer la compréhension, le partage et l’intégration des informations qui lui sont associées. Tout au long de notre travail, nous avons pris en considération les travaux antérieurs et leurs limites. Nous avons utilisé une méthodologie de création d’ontologie qui



Critère VASARI	Définition du critère	Valeurs
F <sub>1</sub> . 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"
F <sub>2</sub> . Lesion side	“Side of lesion epicenter.”	"right", "central", "bilateral"
F <sub>4</sub> . 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"
F <sub>6</sub> . 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"
F <sub>13</sub> . definition of the non enhancing margin	"If most of the outside margin of the non enhancing tumor is well defined and smooth (geographic), versus if the margin is ill-defined and irregular"	"n/a", "irregular", "smooth"
F <sub>20</sub> . Cortical involvement	“Non-enhancing or enhancing tumor extending to the cortical mantle, or cortex is no longer distinguishable relative to subjacent tumor.”	"no", "yes"
F <sub>26</sub> . 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"
F <sub>29</sub> &F <sub>30</sub> . 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"

TABLE 4.5 – Description de quelques critères VASARI [TCGA-Research-Group 2014], la valeur "n/a" ("not applicable") indique que le critère n'est pas applicable au cas étudié et la valeur "none" est attribuée quand l'objet d'étude du critère n'existe pas ou bien n'est pas observable sur la séquence d'image.

est différente de celle de QIBO, nous citerons trois points de différences majeurs : premièrement, nous avons créé IBO autour des concepts qui proviennent d’ontologies spécialisées (FMA, OBI, PATO, etc.) reconnues par la communauté OBO et d’autres comme ONL-DP qui est spécialisée dans la représentation des données d’imagerie médicale. Deuxièmement, IBO a été construite d’une façon modulaire qui facilite sa réutilisation et sa mise à jour potentielle par des futurs utilisateurs par exemple pour extraire des nouveaux sous-ensembles des ontologies externes, par exemple l’outil Ontofox. Enfin, contrairement à QIBO qui n’a pas aligné ses concepts et relations avec une ontologie de haut niveau, IBO est basée sur BFO.

Nous avons modélisé le concept de biomarqueur d’imagerie d’une manière différente de celle qui est proposée dans [Ceusters 2015]. Contrairement à leur proposition, notre proposition articule le concept de biomarqueur avec les aspects de protocole et de processus de mesurage. Nous avons fait ce choix pour nous aligner au mieux avec la définition de base du concept de biomarqueur d’imagerie qui dit qu’un biomarqueur est une caractéristique qui est objectivement évaluée et mesurée. En se référant au rapport de l’ « Institute Of Medecine » et aux travaux de QIBA, le terme « objectively » veut dire d’une manière précise et reproductible. Or, ces deux aspects de précision et de reproductibilité ne peuvent se réaliser que via la définition des protocoles de mesurage. Cette approche diffère de celle de [Ceusters 2015], pour qui l’objectivité est liée aux propriétés intrinsèques de la qualité observée et non pas au processus de mesurage.

Nous pensons avoir mieux explicité ce concept de biomarqueur d’imagerie que ne l’avaient fait les tentatives antérieures, pour les deux raisons suivantes : nous avons ajouté des concepts fondamentaux qui ne sont pas présents dans QIBO, BiomRKRS et le travail de Ceusters et al., par exemple les biomarqueurs qualitatifs, les régions d’intérêt, les protocoles de mesures et les rôles des biomarqueurs d’imagerie, alors que dans [Ceusters 2015], les auteurs introduisent seulement trois sous-catégories de biomarqueurs d’imagerie qu’ils associent entre elles au moyen d’une disjonction. De plus, nous avons articulé notre conception autour des trois axes sémantiques liés au concept de biomarqueur d’imagerie. La limite principale de notre contribution est que, à l’instar des autres travaux antérieurs mentionnés, elle doit être étendue par d’autres classes spécifiques avant son application à un domaine donné de l’imagerie.

## Chapitre 5

# Sémantique des données d'observation en neuro-imagerie selon un point de vue réaliste

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>123</b>
<b>5.2</b>	<b>Contexte et objectif du travail</b>	<b>124</b>
5.2.1	Approche cognitive versus approche réaliste	124
5.2.2	Sémantique des données d'observation en neuro-radiologie	126
5.2.3	Objectif du travail	127
<b>5.3</b>	<b>Matériel et méthode</b>	<b>127</b>
5.3.1	Conception de l'ontologie VASARI	128
5.3.2	Conception du travail expérimental	128
<b>5.4</b>	<b>Résultats</b>	<b>130</b>
5.4.1	Représentation réaliste et ontologique des résultats d'observation VASARI	130
5.4.2	Expérimentations : annotation et exploitation sémantique des données d'observation VASARI	134
<b>5.5</b>	<b>Discussion</b>	<b>142</b>

---

### 5.1 Introduction

Une analyse comparative entre les deux approches de modélisation à savoir l'approche cognitive et réaliste est donnée dans la Section 5.2. Dans cette même section, nous expliquons les raisons qui nous ont menés à adopter une approche réaliste pour la description des données d'observation en neuro-imagerie. La Section 5.3 décrit la méthodologie de notre travail de développement du modèle et d'expérimentation

autour de la base de données REMBRANDT. Dans la Section 5.4, nous décrivons l'ontologie de domaine développée ainsi que l'outil d'annotation qui permet de fournir des définitions formelles aux données en neuro-imagerie. De plus, nous présentons le travail expérimental que nous avons réalisé sur l'ensemble de données REMBRANDT pour démontrer la valeur ajoutée de notre travail par rapport aux formats standard existants, à savoir DICOM SR et le modèle AIM.

## 5.2 Contexte et objectif du travail

Dans ce travail, nous nous focalisons sur l'étude des "*données d'observation*" qui sont exploitées lors du "*processus d'observation ou de mesurage*" dans les études neuro-radiologiques. Associer une sémantique à une donnée d'observation consiste à mettre la donnée en relation avec d'autres entités participant, soit au phénomène observé (par exemple, la température corporelle d'un sujet), soit à un processus d'observation (par exemple, cette même valeur est reliée à une action d'observation, à son observateur, à l'instrument de mesure utilisé et l'instant de la mesure). Comme pour toute donnée des propriétés peuvent être attribuées à la donnée d'observation pour fournir des informations sur sa localisation, son format, sa signification, etc.

Dans le contexte de la pratique clinique, l'interprétation des entités/phénomènes perçus est le fruit d'un processus cognitif réalisé par le médecin spécialiste ce qui indique la nature a priori subjective des résultats de l'observation [Rubin 2014] qui peuvent être dans certains cas erronés [Rector 1991] [Smith 2006a]. Par exemple, deux neuro-radiologues peuvent caractériser différemment une tumeur cérébrale ou bien ils peuvent estimer différemment le volume d'une tumeur s'ils n'utilisent pas les mêmes instruments de mesurage. D'après Smith, une production automatique des résultats d'observation en radiologie est nécessaire d'une part pour réduire cette inter-subjectivité, et d'autre part pour permettre une recherche avancée en imagerie en capturant des observations dans un format standardisé qui supporte le raisonnement basé sur la logique.

En termes de modélisation ontologique, il existe deux approches de modélisation qui sont adoptées pour décrire sémantiquement le contenu sémantique d'une image : *l'approche cognitive* [Cimino 2006] et *l'approche réaliste* [Smith 2006a]. L'approche cognitive oriente sa modélisation autour des "*concepts*" décrits par des "*termes*" faisant partie d'un lexique spécifique que nous manipulons pour l'attribution des propriétés (des qualités ou des dimensions) des données d'observation ; ces termes sont construits selon notre perception et connaissance des entités du monde réel. Contrairement à l'approche cognitive, l'approche réaliste aligne les termes des terminologies aux entités qui existent dans le monde indépendamment d'agents cognitifs reconnaissant leur existence.

### 5.2.1 Approche cognitive versus approche réaliste

Les deux approches cognitive et réaliste proposent des définitions distinctes du terme "concept". Selon Cimino, le terme "concept" est une "unité de traitement

symbolique" de la construction des terminologies médicales. Smith considère les terminologies fondées sur des concepts comme des "collections d'éléments qui peuvent se référer ou non à des entités qui existent dans le monde réel" (par exemple, le concept de diagnostic médical n'existe pas en réalité mais peut être modélisé par des terminologies) et qui "regroupent les termes que les spécialistes utilisent pour exprimer leurs idées". Dans [Smith 2006a], Smith a mentionné les principales limites des terminologies qui sont conçues selon l'approche cognitive, entre autres : 1) la représentation non complète du monde réel, car les concepts ne se réfèrent qu'aux entités du niveau de modélisation abstrait et elles ne précisent pas à quelles instances de ces universels ils font référence, 2) la modélisation non adéquate des entités non observables lors du processus d'observation et 3) l'interprétation confuse des données car il n'existe pas de cadre d'identification cohérent des entités.

En contre partie, Smith distingue dans l'approche réaliste trois niveaux de connaissances [Smith 2006b] :

- Niveau 1 : les objets, les processus, les qualités, les états, etc. du monde réel ;
- Niveau 2 : les représentations cognitives de cette réalité par les scientifiques ou les spécialistes du domaine ;
- Niveau 3 : les concrétisations de ces représentations cognitives dans des artefacts de données.

De plus, il considère qu'il n'y a "qu'une seule réalité objective universelle" ; "chaque attribut du patient est lui-même une entité unique en réalité et on lui attribue son propre identifiant". Ainsi, les entités réelles référencées peuvent être de différents types : entités réelles (l'ensemble de polypes observés au cours d'une coloscopie), mesures (la mesure de la température corporelle d'un patient), etc.

Pour résumer, contrairement à l'approche cognitive, l'approche réaliste ne fait référence qu'aux entités réelles telles qu'elles existent dans le monde. Dans sa vision conceptuelle, Cimino ne décrit pas "ce qu'il existe dans le corps du patient". Par conséquent, l'approche réaliste assure une représentation proche d'une partie de la réalité en se référant explicitement à des instances des entités génériques (universels) et en représentant des interrelations entre elles. Les identificateurs uniques, dans l'approche réaliste sont attribués à des instances d'universaux plutôt qu'à des concepts. Par conséquent, l'identité du même particulier peut être préservée à des moments successifs. Par ailleurs, dans l'approche cognitive, lorsque l'entité change son aspect, un nouveau code est attribué à l'entité référente pour exprimer cette évolution. Ainsi, nous ne pouvons pas générer un historique sur l'évolution d'une entité spécifique. Cette limite de l'approche cognitive rend impossible de suivre les entités selon leur évolution (entités qui n'existent plus ou changement de type, etc.). Cependant, en termes de mise en œuvre, l'approche basée sur les concepts est une solution simple d'annotation de données étant donné qu'elle ne nécessite pas le développement de systèmes particuliers pour générer des identifiants d'entités et gérer des entités complexes.

## 5.2.2 Sémantique des données d'observation en neuro-radiologie

En radiologie, la description sémantique des données d'observation n'est pas une tâche triviale pour deux raisons principales : premièrement, les images médicales avec les données d'observation sont sémantiquement riches et elles font référence aux entités observées (ex : un kyste, un artifact, une lésion, etc.) qui peuvent exister ou non dans le corps du patient [Ceusters 2006]. Deuxièmement, les propriétés décrivant ces entités observées peuvent évoluer dans le temps [Ceusters 2005]. Par exemple, la nature d'une tumeur spécifique peut changer dans le temps (ex : de bénigne à maligne). Ainsi, dans la description de ces données on doit faire référence à la même entité identifiée (ex : lésion) mais de différentes manières (ex : maligne, agrandie, qui disparaît après un certain temps, etc.) pendant toute la durée de vie du patient. Plus spécifiquement, dans le cadre de ce travail les données d'observation considérées sont les résultats des processus d'évaluation des tumeurs cérébrales. Ces données d'observation décrivent l'aspect anatomique, fonctionnel ou moléculaire de la tumeur ou d'une partie de la tumeur. Ces évaluations font référence à des entités qui existent dans le corps du patient et qui sont observables ou non observables sur les séquences d'images.

Jusqu'à aujourd'hui, il n'existe aucun modèle d'information ou ontologique qui permet d'explicitier la définition des connaissances relatives à chaque composante de la tumeur. Telle qu'il est exprimé par Smith : "les terminologies médicales existantes définissent une grande variété de classes abstraites, mais elles ne font référence qu'à un nombre restreint d'instances qui concernent uniquement les êtres humains, l'aspect temporel et les lieux". Par conséquent, la plupart des terminologies médicales ne permettent pas de "suivre une seule et même donnée (par exemple, une tumeur spécifique) sur une période de temps", ni de "distinguer des exemples multiples des mêmes particularités et des particularités multiples du même genre général". Les limites de Radlex, SNOMED, DICOM SR et du modèle AIM vis-à-vis de la description de la sémantique des données d'imagerie sont détaillées dans le Chapitre 3.

La couverture de toutes les informations impliquées dans l'évaluation des tumeurs cérébrales est impossible car aucune source consensuelle n'existe pour spécifier les exigences précises de ce domaine. Pour surmonter cette difficulté, nous avons limité notre étude au domaine couvert par la terminologie VASARI. VASARI constitue un cas d'utilisation représentatif de l'ensemble minimaliste des connaissances basiques qui nécessitent une modélisation formelle.

La terminologie VASARI telle qu'elle est décrite facile à utiliser par les neuro-radiologues, en particulier lorsque des interfaces conviviales sont disponibles. Par exemple, l'outil Clear Canvas<sup>1</sup> qui a été employé par des centres de recherche pour collecter les scores VASARI sous forme d'instances AIM afin de faciliter l'évaluation des annotations (voir Figure 5.1). Cependant, cette méthode d'annotation n'a pas été implémentée sous forme ontologique pour générer les données d'observation radiologique liées aux glioblastomes. Ainsi, ces données n'explicitent pas à quelles

---

1. <https://www.clearcanvas.ca/>

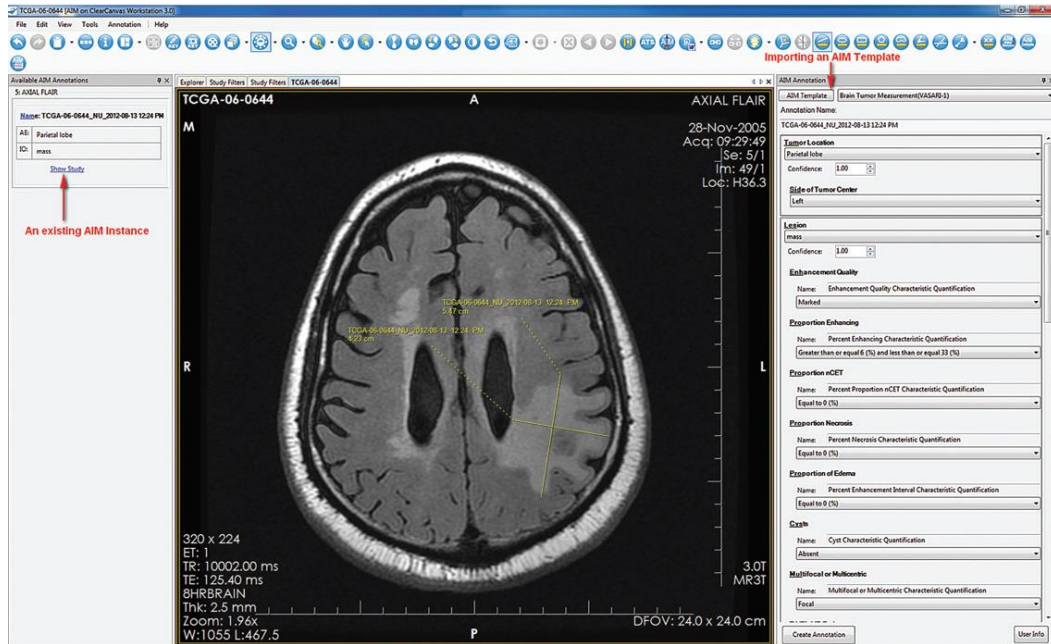


FIGURE 5.1 – La capture d’écran montre un exemple de l’interface Clear Canvas utilisée pour la saisie des données de la terminologie VASARI [Gutman 2013]

entités réelles les neuro-radiologues font référence dans leur évaluation.

### 5.2.3 Objectif du travail

L’objectif principal de ce travail est de faciliter l’identification, le partage et le raisonnement sur les résultats d’observation des tumeurs cérébrales via la formalisation de leurs significations sémantiques. Ce travail va faciliter l’exploitation des connaissances neuro-radiologiques à la fois dans la pratique clinique et la recherche. Nous avons réalisé notre analyse autour de l’étude de la terminologie VASARI comme un cas de preuve de concept. Mais, en réalité notre travail est générique dans la mesure où il reste indépendant de tout domaine de spécialité radiologique et il peut être utile dans d’autres contextes d’imagerie biomédicale (par exemple, la sclérose en plaques, la maladie d’Alzheimer, etc.).

## 5.3 Matériel et méthode

Au cours du processus d’évaluation d’une tumeur cérébrale, le neuro-radiologue attribue une valeur à chaque critère d’imagerie pour attribuer une description standardisée des aspects pertinents de la tumeur qui devraient être pris en compte dans la prise de décision clinique. L’annotation de ces caractéristiques d’imagerie implique différents types d’entités : les entités physiques, les qualités liées aux objets

physiques et les mesures de volume et de taille. Selon la terminologie VASARI, les entités physiques qui caractérisent certaines anomalies des tissus cérébraux sont : la tumeur cérébrale, l'épicentre de la tumeur cérébrale, les composantes de la tumeur cérébrale (à savoir : région de prise de contraste, région de non prise de contraste, partie nécrotique, composante œdémateuse et la bordure de la tumeur cérébrale) et la partie périphérique d'une tumeur cérébrale ou d'une partie de la tumeur cérébrale.

### 5.3.1 Conception de l'ontologie VASARI

Notre méthodologie de modélisation se compose de cinq étapes principales qui peuvent être décrites comme suit : tout d'abord, nous avons analysé la signification de l'aspect étudié par chaque critère VASARI  $F_i$  et nous avons trié ses configurations possibles pour établir la liste des valeurs possibles autorisées pour chaque critère. Deuxièmement, nous avons identifié les principales entités observées qui sont impliquées dans chaque critère. Troisièmement, nous avons défini les entités observées, soit avec des classes d'ontologies existantes soit avec de nouvelles classes ontologiques qui ont été liées. Quatrièmement, nous avons spécifié et défini les axiomes caractérisant ces entités et les relations entre elles. Enfin, nous nous sommes assurés que toutes les configurations possibles pour chaque critère  $F_i$  sont bien modélisées de manière formelle.

Après une analyse de la signification des caractéristiques étudiées par les critères VASARI et de l'identification des différentes entités qu'elles impliquent, nous avons procédé à leur description formelle. Cette étape n'a pas été une tâche très simple pour nous étant donné que nous avons rencontré plusieurs problèmes de modélisation qui sont résumés dans le tableau 5.3.1. Les problèmes de modélisation concernent : les données d'observations négatives (PM1) [Ceusters 2006], les données de représentation de la connaissance spatiale (PM2) [Bennett 2013] et la représentation des entités complexes (PM3).

Nous avons articulé nos concepts autour de la version 2 de l'ontologie BFO, ce qui facilite l'intégration d'ontologies spécialisées issues de la fonderie OBO. En particulier, nous avons réutilisé les ontologies suivantes : FMA, IAO, PATO, OBI, OGMS, UO et l'ontologie RO (Figure 5.2) ; Dans notre travail, nous avons considéré que les relations RO sont intégrées dans l'ontologie BFO 2.

### 5.3.2 Conception du travail expérimental

Dans notre travail expérimental, nous avons développé un logiciel d'annotation sémantique des données VASARI en utilisant l'ontologie VASARI. Ce logiciel permet à l'utilisateur de transformer la description informelle des 30 critères VASARI en une description formelle. Ce logiciel a été appliqué au corpus de données REMBRANDT. L'ensemble de données sémantiques résultant a été utilisé pour évaluer la valeur ajoutée de notre travail. En particulier, nous avons effectué certaines tâches de raisonnement en formulant des requêtes sémantiques ainsi que des tests de cohérence pour détecter des affirmations contradictoires.



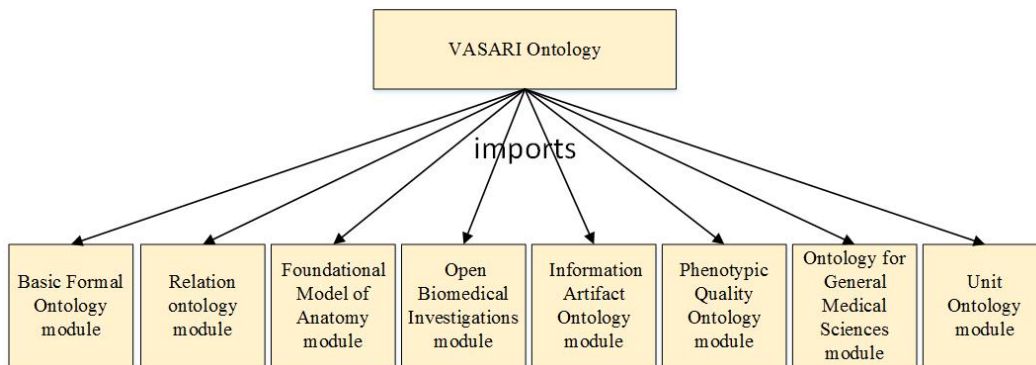


FIGURE 5.2 – La structure générale de l'ontologie VASARI

### 5.3.2.1 Présentation de la base de données REMBRANDT

La base de données REMBRANDT est librement accessible sur ce lien<sup>2</sup>. L'objectif de cette base de données est de faciliter la découverte de corrélations significatives entre les informations cliniques et génomiques afin de fournir aux patients des traitements plus personnalisés dans le contexte clinique. L'ensemble de données REMBRANDT contient 30 critères VASARI annotés par 3 radiologues et relatif à 34 patients atteints de GBM. Les résultats d'observation sont archivés dans un fichier Excel où chaque feuille de calcul contient des évaluations soumises par un radiologue.

### 5.3.2.2 Détails d'implémentation

Nous avons développé l'ontologie VASARI en format OWL2 en utilisant la version 5 de l'outil Protégé. Nous avons utilisé Ontofox pour l'extraction des ontologies OBO. Notre outil d'annotation sémantique est implémenté en langage JAVA (environnement de programmation NetBeans IDE 8.0.2). Nous avons utilisé la version 2.8.3 de l'API JENA [McBride 2002] et la version 2.3.2 du moteur de raisonnement JENA-Pellet (moteur externe de l'API JENA) [Sirin 2007] pour la gestion des données sémantiques (les données d'instances sont sérialisées au format RDF/XML). Pour exécuter certaines requêtes SPARQL, nous avons utilisé le moteur de recherche sémantique CORESE<sup>3</sup> 3.2.

2. <https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project>

3. <http://wimmics.inria.fr/corese>

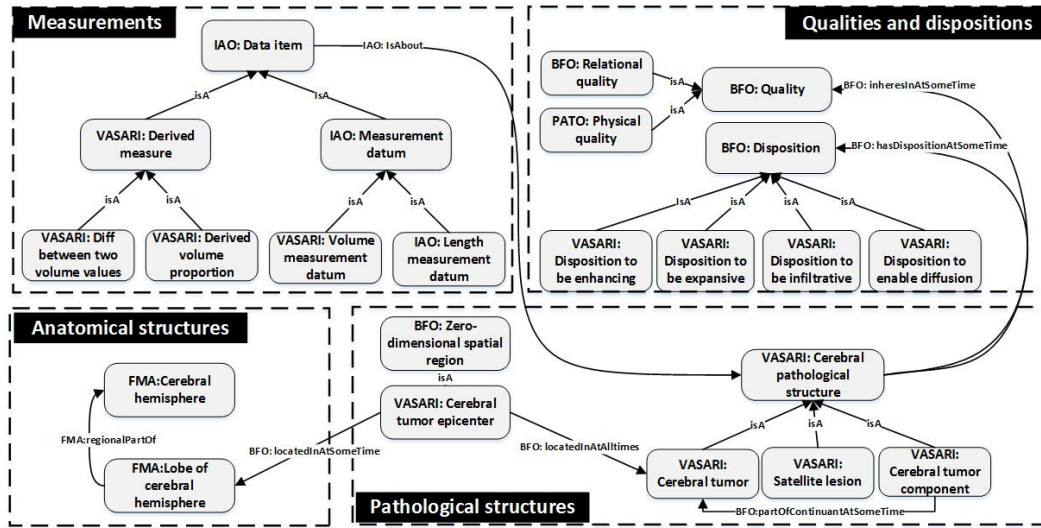


FIGURE 5.3 – Le modèle de base des principales classes dans l'ontologie VASARI

## 5.4 Résultats

### 5.4.1 Représentation réaliste et ontologique des résultats d'observation VASARI

L'ontologie VASARI est composée de huit modules d'ontologies et contient environ 570 classes OWL et 120 propriétés. La figure 5.3 met en évidence les quatre aspects sémantiques majeurs qui décrivent le domaine VASARI, à savoir : les structures pathologiques, la localisation anatomique, les qualités et dispositions, et les mesures.

Dans le modèle sémantique que nous proposons, l'entité "VASARI:cerebral tumor" est définie comme sous classe de l'entité "VASARI:cerebral pathological structure". Les différentes régions de la tumeur cérébrale sont référencées par l'entité "VASARI:cerebral tumor component" et elles sont définies de la façon suivante : "VASARI:cerebral tumor component"  $\equiv$  def. "VASARI:cerebral pathological structure" and "BFO:partOfContinuantAtsomeTime" some "VASARI:cerebral tumor". Nous avons considéré qu'une tumeur cérébrale peut être composée de quatre composants de base qui caractérisent l'anomalie du tissu cérébral : "VASARI:enhancing cerebral tumor component", "VASARI:non enhancing cerebral tumor component", "VASARI:necrotic cerebral tumor component" et "VASARI:cerebral edema component". La figure 5.4 montre la catégorisation complète de sous-classes de l'entité "VASARI:cerebral tumor component".



FIGURE 5.4 – Hiérarchie des classes de type "VASARI:cerebral tumor component" sous Protégé

#### 5.4.1.1 MP1 : conceptualisation des observations négatives

Afin de respecter les principes fondamentaux de l'ontologie réaliste, aucune instance ne doit être créée lorsqu'aucune entité concrète n'existe en réalité. La solution que nous proposons pour la représentation des observations négatives utilise un axiome de classe équivalente (condition 'si et seulement si' impliquant des valeurs négatives de l'assertion). Les deux exemples de définitions de classes montrent comment nous avons introduit la notion d'absence pour les deux catégories  $C_1$  et  $C_2$  :

- Catégorie  $C_1$  : la classe "VASARI:cerebral tumor component not located in brain cortex"  $\equiv$  def. isA "VASARI:cerebral tumor component" and not ("locatedInAtSomeTime" some "FMA:cerebral cortex"), la classe "VASARI:enhancing cerebral tumor without non enhancing cerebral tumor component"  $\equiv$  def. isA "VASARI:enhancing cerebral tumor" and not ("hasContinuantPartAtSomeTime" some "VASARI:non enhancing cerebral tumor component").
- Catégorie  $C_2$  : la classe "VASARI:non cystic cerebral tumor component"  $\equiv$  def. isA "VASARI:cerebral tumor component" et not ("hasQualityAtSomeTime" some "PATO:cystic"), un "VASARI:non enhancing cerebral tumor component"  $\equiv$  def. isA "VASARI:Cerebral tumor component" et not ("hasDispositionAtSomeTime" some "VASARI:disposition to be enhancing").

#### 5.4.1.2 MP2 : conceptualisation des relations entre les entités

Trois principales relations spatiales sont modélisées :

- Cas d'inclusion : Nous avons utilisé la relation spatiale "BFO:locatedInAtSomeTime" et la relation fondamentale "BFO:partOfContinuantAtSomeTime". Nous supposons ici que  $C_1$  et  $C_2$  sont des classes de "BFO:continuant" et que "C1 "BFO:partOfContinuantAtSomeTime" C2" ceci veut dire tel qu'il est exprimé en anglais dans BFO que "for every particular c1, if c1 instanceOf C1 then there is some c2 such that c2 "instanceOf" C2 and c1 "BFO:partOfContinuantAtSomeTime" c2, C1 "locatedInAtSomeTime" C2 asserts that for every c1 if c1 instanceOf C1, then there is some c2 instanceOf C2 and c1 "locatedInAtSomeTime" c2".

Dans notre modèle, nous avons utilisé la relation "BFO:locatedInAtSomeTime", par exemple pour associer l'épicentre de la tumeur "VASARI:cerebral tumor epicenter" à sa localisation "FMA:lobe of cerebral hemisphere" et la relation "BFO:hasContinuantPartAtSomeTime" pour définir que la "VASARI:edematous cerebral tumor" classe  $\equiv$  def. IsA "VASARI:cerebral tumor" et ("hasContinuantPartAtSomeTime" some "VASARI:cerebral edema component").

- Cas de "chevauchement vs adjacence" : Nous avons employé la relation spatiale "RO:adjacentTo" pour exprimer que deux entités de type

"BFO:continuant" ne partagent pas une région spatiale commune et nous avons défini la relation "VASARI:overlaps" pour représenter le cas du chevauchement. Comme décrit dans le tableau 4.5 du Chapitre précédent, F20 évalue l'emplacement de la tumeur cérébrale par rapport au cortex cérébral. Pour décrire ces différentes situations et assurer une classification correcte de la tumeur cérébrale, nous avons défini les classes suivantes : "VASARI: cerebral tumor invading brain cortex"  $\equiv$  def. "VASARI:cerebral tumor" and (VASARI:overlaps some "VASARI:cerebral cortex"), "VASARI:cerebral tumor adjacent to brain cortex"  $\equiv$  def. "VASARI:cerebral tumor" and ("RO:adjacentTo" some "VASARI:cerebral cortex").

- Cas de "séparation" : Nous avons utilisé la qualité relationnelle "VASARI:contiguous with cerebral tumor" avec l'opérateur de négation logique ("not") pour qualifier et identifier un composant cérébral qui est séparé de la tumeur cérébrale, par exemple : "VASARI:satellite lesion"  $\equiv$  def. "VASARI:cerebral pathological structure" et (not ("has quality at some time" some "VASARI:contiguous with cerebral tumor")) et ("VASARI:hasDispositionAtSomeTime" some "VASARI:disposition to be enhancing").

#### 5.4.1.3 MP3 : conceptualisation des entités complexes

La représentation du critère évaluant l'étendue de la résection d'une partie qui prend le contraste de la tumeur cérébrale semble être une donnée simple à représenter, mais en réalité, ce critère fait intervenir plusieurs informations qui ne sont pas explicites dans la définition du critère. Pour modéliser ce genre de critère, nous avons considéré que la classe "VASARI:enhancing cerebral tumor component" ne conservera pas son identité après et avant la chirurgie. Nous pouvons distinguer ainsi deux entités : l'entité "VASARI:enhancing cerebral tumor component before surgery"  $\equiv$  def. IsA "VASARI:enhancing cerebral tumor component" and ("BFO:partOfContinuantAtSomeTime" some "VASARI:enhancing cerebral tumor component") and ("OBI:isSpecifiedInputOf" some "VASARI:resection of enhancing cerebral tumor component"). Et l'entité "VASARI:enhancing cerebral tumor component after surgery"  $\equiv$  def. IsA "VASARI:enhancing cerebral tumor component" and ("BFO:partOfContinuantAtSomeTime" some "VASARI:enhancing cerebral tumor component before surgery") and ("OBI:isSpecifiedOutputOf" some "VASARI:resection of enhancing cerebral tumor component").

Nous avons considéré que la qualité mesurée, c'est-à-dire le volume, est la même, mais les valeurs de mesure du volume sont distinctes : "VASARI:volume measurement datum of enhancing cerebral tumor after surgery"  $\equiv$  def. "VASARI:volume measurement datum" et (isAbout Some "VASARI:enhancing cerebral tumor component after surgery"), "VASARI:volume measurement datum of enhancing cerebral tumor before surgery"  $\equiv$  def. "VASARI:volume

measurement datum" et ("isAbout" Some "VASARI:enhancing cerebral tumor component before surgery"). L'interprétation des valeurs minimales et maximales du critère  $F_{26}$  sera comme suit :

- la proportion 0% signifie que la partie qui prend le contraste dans la tumeur est totalement conservée et que la valeur de son volume avant la chirurgie et après la chirurgie est la même.
- la proportion 100% signifie que la partie qui prend le contraste dans la tumeur est totalement réséquée, de sorte que la composante tumorale cérébrale est classée comme "VASARI:non enhancing cerebral tumor component after surgery".

Le Tableau 5.2 donne des exemples illustratifs pour décrire comment nous avons encodé certains critères VASARI en utilisant notre ontologie. La colonne 1 énumère le critère VASARI étudié, la colonne 2 cite les classes principales qui sont impliquées dans la description des entités observées et la colonne 3 cite les relations principales qui sont utilisées pour relier les instances des différentes classes.

## 5.4.2 Expérimentations : annotation et exploitation sémantique des données d'observation VASARI

### 5.4.2.1 Logiciel d'annotation sémantique des données VASARI

Le logiciel d'annotation prend comme donnée d'entrée l'ensemble des valeurs des critères d'imagerie de la base REMBRANDT ainsi que le schéma de l'ontologie VASARI. Ensuite, pour annoter sémantiquement les données, le logiciel réalise quatre tâches principales : tout d'abord, il crée des instances des classes de l'ontologie VASARI en se basant sur les valeurs des critères. Deuxièmement, il décrit les critères d'imagerie en générant des triplets RDF qui établissent des liens sémantiques entre les instances. Troisièmement, il crée dans un graphe RDF des assertions à partir des triplets. Quatrièmement, il sérialise les données en RDF/XML et enregistre le graphe RDF en mémoire. On note que le logiciel archive séparément le schéma de l'ontologie et les données d'instances (dans les prochains paragraphes, on emploie le terme Tbox pour faire référence au schéma et Abox pour désigner la base d'instances). En terme de performance, le logiciel génère le graphe RDF de l'ensemble de données contenues dans la base REMBRANDT en 1.06 s (soit  $\approx 0.47s$  par feuille).

**Détails sur la génération des instances RDF** Le modèle d'ontologie JENA (une extension du modèle JENA RDF) propose de nombreuses fonctionnalités pour gérer les ontologies. Jena ModelFactory est utilisé pour créer des modèles ontologiques. Le moyen le plus simple de créer un modèle d'ontologie est d'appeler la méthode `createOntologyModel()` de la classe `ModelFactory`. L'instruction est la suivante : `OntModel m = ModelFactory.CreateOntologyModel()`. Dans notre application, nous avons défini un modèle d'ontologie avec les paramètres `OWL_MEM_MICRO_RULE_INF` pour activer l'utilisation le moteur d'inférence OWL.

Pour importer des modules d'ontologies, nous avons utilisé la classe `OntDocumentManager()` qui permet d'établir le lien entre les différents IRI d'ontologies importées et leurs fichiers OWL locaux. Le code présenté dans le Listing 5.1 explique comment nous avons construit le modèle d'ontologie VASARI.

Listing 5.1 – Code to create the VASARI ontology model

```

public OntModel OntModelOntologyRead(String OntologyFileName, String
    owlDirImports, String language)
{
    OntModel ontmodel=null;
    java.io.InputStream in=null;

    try
    {
        ontmodel = ModelFactory.createOntologyModel(OntModelSpec.
            OWL_MEM_MICRO_RULE_INF);
        in=FileManager.get().open(OntologyFileName);
        OntDocumentManager mgr= ontmodel.getDocumentManager();

        logger.info("Importing ontologies' modules");
        mgr.addAltEntry("http://purl.obolibrary.org/obo/obi/2016-06-30/
            OBI_extracted_terms.owl", owlDirImports+"\\OBI_extracted_terms.owl"
        );
        mgr.addAltEntry("http://www.ifomis.org/obo/ro/1.0'', owlDirImports+"\\
            RO.owl");
        mgr.addAltEntry("http://www.semanticweb.org/eamdouni/ontologies/2015/4/
            ONL_DP_extracted_terms", owlDirImports+"\\ONL_extracted_terms.owl");
        mgr.addAltEntry("http://ontofox.hegroup.org/PATO_extracted_terms.owl",
            owlDirImports+"\\PATO_extracted_terms.owl");
        mgr.addAltEntry("http://ontofox.hegroup.org/OGMS_extracted_terms.owl",
            owlDirImports+"\\OGMS_extracted_terms.owl");
        mgr.addAltEntry("http://ontofox.hegroup.org/FMA_extracted_terms.owl",
            owlDirImports+"\\FMA_extracted_terms.owl");
        mgr.addAltEntry("http://purl.obolibrary.org/obo/bfo.owl", owlDirImports
            +"\\BFO.owl");
        mgr.addAltEntry("http://purl.obolibrary.org/obo/iao/2015-02-23/
            IAO_extracted_terms.owl", owlDirImports+ "\\IAO_extracted_terms.owl
        ");
    }

    catch (Exception e)
    {
        e.getMessage();
    }

    if (in==null)
    {
        throw new IllegalArgumentException("File not found");
    }
    else
    {

```

```
logger.info("Generation of the VASARI ontmodel");  
return (OntModel) ontmodel.read(in, null);  
}  
}
```

Le code fourni dans le Listing 5.2 montre un exemple de création d'une assertion à partir du schéma VASARI et d'enregistrement des assertions créées dans un modèle de données RDF. Pour réaliser ces fonctions, le logiciel devrait définir et créer les structures suivantes :

- les Ontclasses via la méthode `getOntClass()`;
- les individus via la méthode `createIndividual()`;
- les propriétés objet/ données via les méthodes `createObjectProperty()/ createDataProperty()`, respectivement.

Nous nous référons à `OntClasses` à partir du modèle d'ontologie (introduit dans le code sous le nom de `vasariOntology`) et nous ajoutons des individus et des déclarations dans le modèle qui contient les données RDF (introduit dans le code sous le nom de `instanceModel`). Deuxièmement, les propriétés entre les instances sont ajoutées faisant appel à la méthode `createObjectProperty()`. Enfin, nous ajoutons ces axiomes au modèle d'instance comme des assertions.

Listing 5.2 – Un exemple de génération de données d'instance à partir du modèle d'ontologie VASARI

```
public void RDFDataCreation (VASARISheetRow row, OntModel  
    vasariOntology, OntModel instanceModel)  
{  
    OntClass cerebralTumorEpicenterClass, cerebralTumorClass;  
    Individual cerebralTumorEpicenterIndiv, cerebralTumorIndiv;  
    Property located_in_at_some_time;  
  
    cerebralTumorEpicenterClass = vasariOntology.getOntClass(vasariOntology  
        .getNsPrefixURI("VASARI_NS") + "cerebral_tumor_epicenter");  
    cerebralTumorEpicenterIndiv=instanceModel.createIndividual(  
        instanceModel.getNsPrefixURI("VASARI_NS") + "_cte_" + row.  
        getPatientNum(), cerebralTumorEpicenterClass);  
  
    cerebralTumorClass= vasariOntology.getOntClass(vasariOntology.  
        getNsPrefixURI("VASARI_NS")+ "cerebral_tumor");  
    cerebralTumorIndiv=instanceModel.createIndividual(instanceModel.  
        getNsPrefixURI("VASARI_NS") + "_ct_" + row.getPatientNum(),  
        cerebralTumorClass);  
  
    located_in_at_some_time = instanceModel.createObjectProperty(  
        vasariOntology.getNsPrefixURI("OBO_NS")+ "BFO_0000171");  
  
    cerebralTumorEpicenterIndiv.addProperty(located_in_at_some_time,  
        cerebralTumorIndiv);  
    Statement s = instanceModel.createStatement(cerebralTumorEpicenterIndiv  
        , located_in_at_some_time, cerebralTumorIndiv);  
    instanceModel.add(s);  
}
```



}

Nous avons stocké les données d'instance générées dans un fichier RDF/XML. Par conséquent, trois sortes des fichiers RDF/XML ont été créés :

- Un fichier pour sauvegarder les observations d'un seul neuro-radiologue pour un patient donné ;
- Un fichier pour enregistrer le résultat d'agrégation des observations des trois neuro-radiologues pour un patient donné ;
- Un fichier pour enregistrer le résultat d'agrégation des observations faites par un seul neuro-radiologue pour l'ensemble de patients.

Pour écrire les déclarations générées qui sont contenues dans le modèle RDF, nous avons utilisé la méthode `write()` du modèle de classe et nous avons spécifié le format "RDF/XML" comme un langage de sérialisation.

Le Listing 5.3 montre un ensemble de triplets RDF générés après l'annotation des données d'observation du patient 900\_00\_1961 produites par le neuro-radiologue 1 : "tumor location= 2", "lesion side=1", "eloquent brain= 1", "enhancement quality=1", "proportion enhancing= 7", "proportion non enhancing=1", "proportion necrosis= 1", "cyst= 1", "multicentric=1", "T1/FLAIR= 1", "thickness of enhancing margin= 1", "definition of the enhancing margin= 1", "definition of non enhancing margin=2", "proportion of edema= 3", "edema crosses midline=2", "hemorrhage= 1", "diffusion=2", "pial invasion= 1", "ependymal invasion=1", "cortical involvement= 2", "deep white matter invasion= 1", "nCET tumor crosses midline= 2", "enhancing tumor crosses midline= 1", "satellites= 1", "calvarial remodeling=1", "extent of resection of enhancing tumor=1", "extent of resection of non enhancing tumor=6", "extent of resection of vasogenic edema= 2", "lesion size=(12,7)".

Listing 5.3 – Annotation des données d'observation du patient 900\_00\_1961 produites par le radiologue 1

```

<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:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:VASARI_NS="http://www.semanticweb.org/eamdouni/ontologies/2016/1/
VASARI_ontology#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
<owl:ObjectProperty rdf:about="http://www.semanticweb.org/eamdouni/
ontologies/2016/1/VASARI_ontology#overlaps"/>
<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
BFO_0000171"/>
<owl:ObjectProperty rdf:about="http://www.ifomis.org/obo/ro/1.0#
adjacentTo"/>
<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
IAO_0000221"/>
<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
BFO_0000112"/>

```

```

<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
  IAO_0000136"/>
<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
  BFO_0000086"/>
<owl:ObjectProperty rdf:about="http://purl.obolibrary.org/obo/
  BFO_0000176"/>
<owl:DatatypeProperty rdf:about="http://purl.obolibrary.org/obo/
  IAO_0000004"/>
<VASARI_NS:cerebral_tumor_epicenter_not_located_in_eloquent_cortex rdf:
  about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/
  VASARI_ontology#_cte_900_00_1961">
<OBO_NS:BFO_0000171>
<OBO_NS:FMA_72971 rdf:about="http://purl.obolibrary.org/obo/
  _rtl_900_00_1961"/>
</OBO_NS:BFO_0000171>
<OBO_NS:BFO_0000171>
<VASARI_NS:cerebral_tumor_not_invading_ependyma rdf:about="http://www.
  semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#
  _ct_900_00_1961">
<VASARI_NS:overlaps>
<OBO_NS:FMA_61830 rdf:about="http://purl.obolibrary.org/obo/
  _cc_ct_900_00_1961"/>
</VASARI_NS:overlaps>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#cerebral_tumor_not_invading_pia_mater"/>
<OBO_NS:BFO_0000112>
<VASARI_NS:disposition_to_facilitate_diffusion rdf:about="http://www.
  semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#
  _dtfd_ct_900_00_1961"/>
</OBO_NS:BFO_0000112>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#non_hemorrhagic_cerebral_tumor"/>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#non_cystic_cerebral_tumor"/>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#non_enhancing_cerebral_tumor"/>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#necrotic_cerebral_tumor"/>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#cerebral_tumor"/>
</VASARI_NS:cerebral_tumor_not_invading_ependyma>
</OBO_NS:BFO_0000171>
<rdf:type rdf:resource="http://www.semanticweb.org/eamdouni/ontologies
  /2016/1/VASARI_ontology#cerebral_tumor_epicenter"/>
</VASARI_NS:cerebral_tumor_epicenter_not_located_in_eloquent_cortex>
<VASARI_NS:cerebral_tumor_component rdf:about="http://www.semanticweb.
  org/eamdouni/ontologies/2016/1/VASARI_ontology#_ctc_900_00_1961">
<OBO_NS:BFO_0000112>
<VASARI_NS:disposition_to_be_expansive rdf:about="http://www.
  semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#
  _dtbex_ctc_900_00_1961"/>
</OBO_NS:BFO_0000112>
<OBO_NS:BFO_0000176 rdf:resource="http://www.semanticweb.org/eamdouni/
  ontologies/2016/1/VASARI_ontology#_ct_900_00_1961"/>

```

```

</VASARI_NS:cerebral_tumor_component>
<VASARI_NS:cerebral_edema_component rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_cec_900_00_1961">
<OBO_NS:BFO_0000176 rdf:resource="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_ct_900_00_1961"/>
</VASARI_NS:cerebral_edema_component>
<VASARI_NS:Brain_without_satellite_lesion rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_brain_900_00_1961"/>
<VASARI_NS:eroded_inner_table_of_calvaria rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_itoc_900_00_1961"/>
<VASARI_NS:cerebral_tumor_margin rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_ctm_900_00_1961">
<OBO_NS:BFO_0000176 rdf:resource="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_ct_900_00_1961"/>
</VASARI_NS:cerebral_tumor_margin>
<VASARI_NS:outside_margin_of_enhancing_cerebral_tumor_component rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_omoectc_900_00_1961"/>
<VASARI_NS:necrotic_cerebral_tumor_component rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_nctc_900_00_1961">
<OBO_NS:BFO_0000176 rdf:resource="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_ct_900_00_1961"/>
</VASARI_NS:necrotic_cerebral_tumor_component>
<VASARI_NS:
  perpendicular_axis_to_the_longest_dimension_measurement_datum 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_0000221>
<VASARI_NS:perpendicular_to_the_longest_axis rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_ptotla_900_00_1961"/>
</OBO_NS:IAO_0000221>
</VASARI_NS:
  perpendicular_axis_to_the_longest_dimension_measurement_datum>
<VASARI_NS:long_axis_measurement_datum rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_lamd_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_0000221>
<VASARI_NS:longest_axis rdf:about="http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#_la_900_00_1961"/>
</OBO_NS:IAO_0000221>
</VASARI_NS:long_axis_measurement_datum>
<OBO_NS:FMA_50801 rdf:about="http://purl.obolibrary.org/obo/_brain_900_00_1961"/>
</rdf:RDF>

```

### 5.4.2.2 Exploitation sémantique des données d'annotation VASARI

Notre application utilise un modèle inféré généré par un raisonneur pour réaliser le raisonnement. Les informations archivées dans le graphe inféré sont contenues dans une base de connaissances (KB) qui peut être exploitée de deux façons : 1) accessible via les requêtes SPARQL pour récupérer les données en fonction de leur sémantique 2) vérifiée via des tests de cohérence.

**Requêtage de la base de connaissances** Considérons cet exemple pour illustrer les capacités d'inférence et démontrer comment le format sémantique permet d'exploiter des connaissances anatomiques provenant de l'ontologie FMA. Nous considérons que la base de connaissances est composée d'un Abox contenant les assertions suivantes : (A1) `cte instanceOf "VASARI:cerebral tumor epicenter"`, (A2) `tl instanceOf "FMA:temporal lobe"`, (A3) `rch "instanceOf" "FMA:right cerebral hemisphere"`, (A4) `cte "locatedInAtSomeTime" tl`, (A5) `tl "regionalPartOf" rch`. A Tbox that contains : (T1) `"FMA:right temporal lobe" ≡ def. "FMA:temporal lobe" and regionalPartOf some "FMA:right cerebral hemisphere"`. En se basant sur les assertions du Abox et l'axiome (T1) on peut déduire : l'assertion (A6) `tl instanceOf "FMA:right temporal lobe"` (voir Figure 5.5) ainsi nous pouvons demander au raisonneur de ressortir uniquement les épïcètres de tumeurs cérébrales qui sont situés dans le lobe temporal droit.

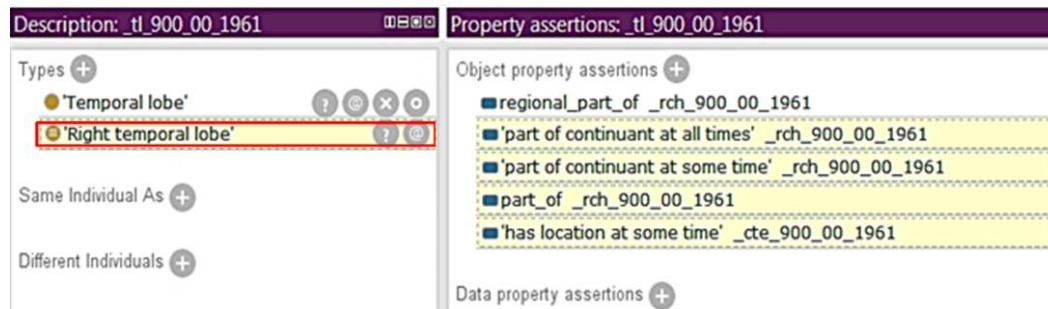


FIGURE 5.5 – Exemple illustratif d'exploitation sémantique des connaissances anatomiques de l'ontologie FMA (sous l'outil protégé) ; les noms des instances sont générés automatiquement avec notre outil d'annotation, l'ID du patient 900\_00\_1961 est inclus pour désigner les entités existantes dans le corps du patient.

La figure 5.6 présente un exemple de requête SPARQL qui permet d'afficher l'emplacement des tumeurs cérébrales (`?lobe`, `?cerebral_hemisphere`) et leurs mesures de longueur correspondantes (`?long_axis_value`, `?perpendicular_long_axis_value`). La partie inférieure de la figure présente les résultats retournés par l'outil de requêtage sémantique CORESE.

**Validation de la base de connaissances** Le raisonnement avec JENA-Pellet permet de détecter les conflits contenus dans la base de connaissances ; nous avons donc exploité cette capacité pour effectuer une vérification globale dans la KB et rechercher les incohérences dans les assertions des radiologues. Par exemple, supposons que (A1) est `instanceOf "VASARI:cerebral tumor"` et qu'un radiologue a déclaré que (A2) est `isA "VASARI:hemorrhagic cerebral tumor"` cependant, un autre radiologue a déclaré que (A3) est `isA "VASARI:non hemorrhagic cerebral tumor"`. Le raisonnement sur cette base de connaissances est impossible étant donné que les deux classes `"VASARI:hemorrhagic cerebral tumor"` et `"VASARI:non hemorrhagic cerebral tumor"` sont définies comme des classes disjointes dans la Tbox ; cela signifie qu'ils ne peuvent pas partager le même ensemble d'instances. Ceci devrait se traduire par une erreur de classification détectée par le raisonneur.

**Détails sur la validation du modèle** Le résultat d'une vérification de cohérence est fourni via l'objet `"ValidityReport"` de l'API JENA. Cette structure encapsule tous les axiomes et assertions inconsistants qui sont détectés. Pour générer des explications sur les causes d'incohérences, nous avons utilisé la méthode `"explainconsistency()"`. Cette méthode énumère tous les axiomes impliqués ; la Figure 5.7 représente un exemple d'incohérence qui est causée par l'attribution de deux valeurs distinctes de la mesure de la longueur d'une même tumeur cérébrale. Cette attribution multiple viole l'axiome owl `:functionalProperty` de la propriété `"IAO:has_measurement_value"` qui est déclarée comme fonctionnelle.

Listing 5.4 – code pour valider le modèle inféré

```
public void OntologyConsistencyTest (PelletInfGraph infgraphmodel,
    InfModel infmodel)
{
    Boolean consistencyTest= true;
    logger.info("Generation of the validity Report" + "\n");
    ValidityReport validity = infmodel.validate();
    testConsistency= validity.isValid();
    java.io.OutputStream out=null;

    if (consistencyTest)
    {
        logger.info("The model is consistent");
        JTextArea1.append("The model is consistent");
    }
    else
    {
        logger.info("The model is not consistent: some conflicts are detected:"
            + "\n");
        JTextArea1.append("The model is not consistent: some conflicts are
            detected:" + "\n");
    }

    for (Iterator<ValidityReport.Report> iterator = validity.getReports();
```

```
        iterator.hasNext(); )
    {
    final ValidityReport.Report report = iterator.next();
    jTextArea1.append( "-" + report.getDescription() + "\n" );
    }

    Model explanation = infgraphmodel.explainInconsistency();
    System.out.println("Explanation of inconsistency:");
    explanation.write(System.out, "RDF/XML");
    }
}
```

## 5.5 Discussion

Dans ce travail, nous avons adopté l'approche réaliste pour décrire les principales entités observées dans les tumeurs cérébrales. On note que l'approche réaliste est de plus en plus adoptée par la communauté biomédicale. L'adoption de l'approche réaliste nous a permis de fournir une représentation plus explicite des critères d'imagerie en considérant à la fois le "niveau universel" (par exemple, le concept de tumeur cérébrale) et des "instances" (par exemple, la tumeur cérébrale de David). Pour suivre cette perspective de modélisation, nous avons utilisé l'ontologie réaliste BFO, à savoir l'ontologie BFO pour décrire les entités "existantes" et les relations entre elles. L'utilisation de l'ontologie BFO a facilité l'intégration de connaissances hétérogènes à partir de différentes ontologies spécialisées dans l'anatomie, les qualités, les mesures, etc.

L'ontologie développée répond à certains points problématiques qui sont discutés dans [Ceusters 2006] [Cimino 2006] [Smith 2006b] [Levy 2012], principalement : 1) la formalisation de la description de l'information dans le domaine de la neuro-imagerie par l'utilisation d'ontologies spécialisées ; on peut citer l'exemple de l'ontologie FMA qui a permis la description de certaines assertions cliniques ou l'ontologie BFO qui nous a aidé dans l'établissement de relations fondamentales (c'est-à-dire **partOf**) et spatiales (c.-à-d. emplacement, **locatedIn**, etc.) entre les structures pathologiques, 2) la représentation d'observations négatives en neuro-imagerie par l'utilisation d'axiomes OWL et 3) la représentation d'entités complexes, qui a été plus explicite en se référant aux entités concrètes impliquées. Cependant, nous avons rencontré quelques difficultés pour représenter les expressions mathématiques avec l'ontologie IAO, nous pensons qu'il serait intéressant d'étendre IAO pour couvrir ce type d'information qui est nécessaire dans la définition des expériences biomédicales. Nous avons utilisé la terminologie VASARI comme preuve de concept pour la démonstration de la faisabilité et l'importance de décrire les observations relatives aux tumeurs cérébrales, sous la forme d'instances et de relations entre celles-ci exprimées respectivement sous forme d'instances de classes OWL et relations exprimées sous forme de triplets RDF. Notre travail peut être facilement étendu pour répondre à d'autres cas d'utilisation, grâce à l'aspect modulaire et extensible de l'ontologie et au langage OWL auto-descriptif (les concepts sont textuellement et formellement

décrits dans l'ontologie pour guider les utilisateurs). Il est important de noter que notre approche n'est pas limitée à la description sémantique des critères VASARI résultant de l'évaluation subjective des images IRM par des neuro-radiologues mais peut être appliquée à d'autres spécialités d'imagerie. En revanche, notre travail aurait exprimé sa pleine valeur s'il avait été mis en œuvre en complément d'un système automatisé ou semi automatisé d'analyse d'images [Velazquez 2015]. Par exemple, les systèmes décrits dans ces documents [Rubin 2014] [Porz 2014] sont capables de segmenter les différentes parties de la tumeur et de déterminer automatiquement leur environnement anatomique (par exemple, dans quelles structures anatomiques elles sont situées ou bien si elles se chevauchent ou sont des composantes adjacentes). De telles propriétés méreotopologiques pourraient être traduites directement sous forme sémantique en utilisant les relations discutées ci-dessus. De même, les mesures du volume et les proportions dérivées pourraient être générées automatiquement et avec une meilleure précision que par les évaluations subjectives du modèle VASARI.

Les logiciels d'archivage et de gestion des données médicales ainsi que les systèmes de communication tendent à assister les praticiens dans l'interprétation des images médicales en leur facilitant les tâches d'identification, segmentation et extraction des observations radiologiques à partir des ROIs. Ceci nécessite le développement et la mise en place d'algorithmes pour permettre l'automatisation de ces tâches. De cette façon, si des études antérieures d'un patient spécifique sont disponibles, les régions d'intérêt précédentes seront automatiquement identifiées par le logiciel de lecture. Ainsi, les caractéristiques de taille, de forme, d'emplacement, de texture de la lésion seront disponibles et pourront donc être comparées aux nouvelles valeurs, permettant ainsi des diagnostics plus précis.

La première limite de notre proposition est qu'elle est implémentée en OWL et qu'elle ne génère donc que des instances qui ne prennent pas en compte de l'aspect temporel [Smith 2006b]. Nous pensons que la prise en compte de l'aspect temporel dans la représentation des critères d'imagerie est nécessaire en particulier dans le cadre des études de suivi longitudinal, par exemple pour évaluer la réponse thérapeutique des cellules cancéreuses. Dans ce contexte, nous recommandons de sélectionner un langage logique capable de représenter les relations "n-aire". La deuxième limite est intrinsèque au problème de la contradiction logique qui est due au fait que les radiologues décrivent ce qu'ils observent en fonction de leurs connaissances du domaine. En conséquence, ils peuvent décrire différemment la "réalité" et produire des enregistrements cliniques différents sur les entités identifiées. Pour résoudre le désaccord dans l'interprétation, de nombreux systèmes médicaux nécessitent de conserver dans l'entrée de données une seule valeur "correcte" pour la fonctionnalité évaluée pour faciliter l'agrégation des données. Selon Rector et al., les dossiers médicaux électroniques devraient permettre la présence de déclarations contradictoires, de mesures multiples, etc. pour mieux s'approcher de la réalité.

En résumé, nous pensons que les données de la neuro-imagerie devraient être codées dans un format structuré qui rend leurs significations explicites aux systèmes et facilitent ainsi leur compréhension ainsi que leur gestion. La prise en considération de la sémantique des critères d'imagerie (valeurs de mesure, qualités, composantes de

la tumeur, localisation des lésions, etc.) est nécessaire pour deux raisons principales : (1) améliorer la qualité des soins cliniques qui ont tendance à fournir des traitements personnalisés aux patients par l'utilisation de recommandations cliniques qui sont basées sur les critères d'évaluation (2) soutenir la recherche clinique sur le développement de nouveaux biomarqueurs d'imagerie en combinant les données cliniques avec des informations provenant de différents domaines médicaux.



Problème de modélisation	Expressions VA-SARI	Exemples de cas d'utilisation
PM1 : Comment représenter des observations négatives de neuroimagerie qui indiquent la non-existence d'un "BF0:dependant continuant" (catégorie C <sub>1</sub> ) ou d'un "BF0:independant continuant" (catégorie C <sub>2</sub> ) : "BF0:quality" ou "BF0:disposition" ?	Les résultats négatifs sont exprimés par l'utilisation d'expressions de négation comme par exemple "none", "indeterminate", "not applicable", "without", etc.	La non existence d'une entité ou d'une qualité/-disposition. Exemples : C <sub>1</sub> : "La tumeur cérébrale de John est sans région de prise de contraste". C <sub>2</sub> : "La tumeur cérébrale de John n'est pas érodée". ou bien "la tumeur cérébrale de John n'est pas infiltrante".
PM2 : Comment réaliser une représentation proche de la réalité de la structure pathologique étudiée et décrire comment ses composants sont situés dans l'espace ?	Localisation géographique des entités existantes, entités liées l'une à l'autre, entités séparées les unes des autres, entités adjacentes, etc.	Exemples : "L'épicentre de la tumeur cérébrale de John est situé dans le lobe pariétal". "La tumeur cérébrale a une marge".
PM3 : Comment coder des entités complexes comme par exemple des mesures dérivées (proportions de mesures de volume, mesures de longueur, etc.) et les associer à leurs résultats cliniques correspondants	La représentation de l'étendue de la résection d'un composant tumoral cérébral donné (nCET, nécrotique, etc.).	Exemples : "35% de la tumeur cérébrale de John prend le contraste", "la taille de la tumeur cérébrale de John est de 1 cm x 2 cm".

TABLE 5.1 – Résumé de la nature des problèmes de modélisation rencontrés dans la description sémantique des critères VASARI

Critère VA-SARI	Certaines entités directement impliquées	Certaines propriétés impliquées
F1 : lesion location	"VASARI: cerebral tumor epicenter", "FMA:lobe of cerebral hemisphere", "FMA:brainstem", FMA:cerebellum	"OBO:Located in at some time"
F2 : lesion side	"VASARI:cerebral tumor epicenter", "FMA:right/left lobe of cerebral hemisphere", "FMA:brainstem", "FMA:cerebellum", "FMA:median sagittal plane"	"OBO:located in at some time"
F3 : eloquent brain	"FMA:primary motor cortex", "FMA:occipital lobe", "FMA:broca's area", "FMA: Wernicke's area"	"OBO:located in at some time"
F4 : enhancement quality	"VASARI:disposition to be enhancing", "VASARI:cerebral tumor", "VASARI:cerebral tumor component"	"OBO:has disposition at some time"
F5 : proportion enhancing	"PATO:volume", "VASARI:cerebral tumor component", "OBI:measurement datum", "OBI:value specification"	"OBO:has quality at some time", "OBO:has disposition at some time"
F24 : satellites	"VASARI:satellite lesion", "FMA:brain"	"OBO:has quality at some time", "OBO:has disposition at some time"
F30 : lesion size	"OBI:length measurement datum", "UO:length unit", "PATO:quality"	"OBO:is quality measurement of", "OBO:has measurement unit label"

TABLE 5.2 – Description des critères VASARI sous forme d'entités et de relations ontologiques

The screenshot shows a SPARQL query execution interface. The query is as follows:

```

PREFIX fma: <http://purl.obolibrary.org/obo/fma#>
PREFIX vasari: <http://www.semanticweb.org/eamdouni/ontologies/2016/1/VASARI_ontology#>
SELECT ?cerebral_tumor ?lobe ?hemisphere ?long_axis_value ?perpendicular_long_axis_value
WHERE {
  ?lobe_class rdfs:subClassOf* obo:FMA_61823
  ?lobe_instance rdf:type ?lobe_class
  ?hemisphere_class rdfs:subClassOf obo:FMA_61817
  ?hemisphere_instance rdf:type ?hemisphere_class
  ?cerebral_tumor_epicenter_instance rdf:type vasari:cerebral_tumor_epicenter
  ?cerebral_tumor_epicenter_instance obo:BFO_0000171 ?lobe_instance
  ?cerebral_tumor_instance rdf:type vasari:cerebral_tumor
  ?cerebral_tumor_instance rdfs:label ?cerebral_tumor
  ?cerebral_tumor_epicenter_instance obo:BFO_0000171 ?cerebral_tumor_instance
  ?cerebral_tumor_epicenter_instance rdfs:label ?cerebral_tumor_epicenter
  ?lobe_instance fma:regional part of ?hemisphere_instance
}

```

The results are displayed in a table with the following columns: ?cerebral\_tumor, ?lobe, ?hemisphere, ?long\_axis\_value, and ?perpendicular\_long\_axis\_value. The table contains 30 rows of data.

?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 5.6 – Résultat d'exécution d'une requête SPARQL sur les observations du neuro-radiologue 1 de la base REMBRANDT (démonstration avec l'outil CO-RESE). FMA\_61823 désigne "FMA:lobe of cerebral hemisphere", FMA\_61817 désigne le "FMA:cerebral hemisphere" et BFO\_0000171 indique la relation "OBI:located\_in\_at\_some\_time".

```
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 5.7 – Exemple d'une sortie de validation : détection d'une contradiction logique, IAO\_000004 désigne la relation "IAO:has\_measurement\_value" et \_paldmd\_900\_00\_1961 fait référence à la valeur de mesure de l'axe perpendiculaire le plus long du patient 900\_00\_1961.

# Chapitre 6

## Discussion

### Contents

---

<b>6.1</b>	<b>Introduction</b>	<b>149</b>
<b>6.2</b>	<b>Rappel des objectifs et contributions de la thèse</b>	<b>150</b>
<b>6.3</b>	<b>Évaluation de nos contributions</b>	<b>150</b>
6.3.1	Évaluation de l'ontologie IBO	150
6.3.2	Évaluation de l'ontologie VASARI et des expérimentations	151
<b>6.4</b>	<b>Évaluation du travail par rapport à l'existant</b>	<b>153</b>

---

### 6.1 Introduction

Actuellement, en routine clinique les mesures cliniques sont rarement incluses dans des rapports structurés, les radiologues préfèrent encore la description textuelle de ce qu'ils observent et quantifient. Ainsi, les résultats d'imagerie obtenus sont difficilement réutilisables dans d'autres contextes cliniques ou de recherche ; ceci est dû à d'une part, la variabilité des paramètres d'acquisition, des procédures de traitement et des méthodes de quantification des paramètres d'imagerie et d'autre part à la subjectivité dans l'interprétation des images médicales. En recherche, les bio-banques d'images existantes stockent les données sans expliciter les connaissances qu'elles décrivent. La mise en place des systèmes d'interrogation sémantiques favorisera certainement les avancées dans la recherche biomédicale, en offrant à la communauté des services qui permettent de faire des requêtes avancées dans le cadre des études multi-centriques.

L'utilisation des ontologies aide à surmonter les problèmes cités ci-dessus en fournissant une définition précise pour chaque entité et relation. Les mécanismes de raisonnement en web sémantique vont aider à faire : la vérification automatique de la cohérence, la classification automatique des concepts, la recherche avancée des données, et l'interrogation des sources hétérogènes et réparties.

## 6.2 Rappel des objectifs et contributions de la thèse

Le sujet de thèse visait un double objectif : (1) proposer un modèle générique de représentation des biomarqueurs d'imagerie, indépendant des modalités d'imagerie, des spécialités médicales et des applications susceptibles de les utiliser ; (2) valider ce modèle en le spécialisant dans un ou quelques domaines importants de la pathologie, et en l'implémentant dans une application capable de mettre en lumière la valeur ajoutée apportée par cette représentation sémantique.

Concrètement, notre travail s'est décliné en deux grandes parties : la première partie concerne le développement de l'ontologie générique IBO pour la formalisation de la description des biomarqueurs d'imagerie. La seconde partie décrit une ontologie de domaine que nous avons conçue selon une approche réaliste pour la description des données d'observations en neuro-imagerie et montre l'intérêt du travail autour de la base REMBRANDT sur la description des glioblastomes.

## 6.3 Évaluation de nos contributions

### 6.3.1 Évaluation de l'ontologie IBO

Nous pensons que nous avons proposé une modélisation "rigoureuse" de la sémantique des concepts génériques des données d'imagerie dans la recherche touchant aux biomarqueurs d'imagerie. Le choix de l'ontologie BFO pour l'articulation des concepts sous différentes catégories (entités physiques, qualités, processus, etc.) facilite la distinction entre les traitements exécutés pour le mesurage de la qualité biologique, les entrées et les sorties du processus de mesurage et les décisions cliniques liées au biomarqueur en question. Nous considérons qu'il est d'une grande importance de faire le lien entre ces principaux aspects fondamentaux des biomarqueurs d'imagerie pour garantir une meilleure compréhension des données de provenance des résultats et ainsi pouvoir les réutiliser correctement dans des études de comparaison, de validation ou autre.

Dans le domaine de la représentation des connaissances "la meilleure" ontologie n'existe pas ; ainsi plusieurs ontologies peuvent être proposées pour répondre à une problématique de modélisation donnée dans un domaine particulier. De notre point de vue, les ontologies existantes définissent la terminologie relative aux biomarqueurs d'imagerie mais aucune d'entre elles n'a levé l'ambiguïté sur la définition de ce concept en focalisant la modélisation sur ces trois questions :

- Quels sont les traitements appliqués pour le calcul d'une certaine valeur d'un biomarqueur d'imagerie ?
- Quel est le rôle clinique et la maladie associés à ce type de biomarqueur d'imagerie ?
- Quelle est la réalité biologique qui est estimée à travers le biomarqueur d'imagerie en question ?

Nous pensons qu'il reste quelques aspects qu'il faut affiner ou réviser dans l'ontologie IBO, particulièrement les qualités issues de l'ontologie PATO et la spécification

des paramètres d'imagerie liées aux protocoles du processus de mesurage des biomarqueurs d'imagerie comme par exemple "subject positioning", "respiratory motion minimization", "scanning direction", etc. qui existent dans le profil "QIBA FDG PET/CT". Néanmoins, nous pensons qu'IBO apporte une bonne base générique de modélisation des principaux aspects relatifs aux biomarqueurs d'imagerie.

Parmi les objectifs que nous avons cités précédemment nous n'avons pas pu spécialiser notre modèle dans un domaine médical particulier comme nous anticipons de le faire parce que nous n'avons pas eu accès à des données d'imagerie dans le cadre du projet ANR ISIMED spécialisation qui aurait permis de valider notre modèle générique au moins partiellement nous sommes malheureusement se contentés de faire quelques illustrations pour montrer comment IBO pourrait être utilisé pour implémenter des vraies valeurs quantitatives et qualitatives des biomarqueurs d'imagerie. Ces illustrations ont montré que l'annotation sémantique des biomarqueurs d'imagerie avec IBO permet de référencer les instances (organes, pathologies, processus décisionnels, etc.) relatives à chaque mesure radiologique.

Ce qui reste à faire pour compléter notre travail est clairement l'exploitation de l'ontologie IBO pour l'implémentation de quelques biomarqueurs d'imagerie dans un cadre applicatif donné. Ceci n'est pas une étape très triviale car elle demande d'une part la collaboration avec des centres radiologiques et d'autre part l'assistance d'experts médicaux pour l'évaluation de la justesse des résultats d'annotations.

### 6.3.2 Évaluation de l'ontologie VASARI et des expérimentations

L'objectif de ce travail était d'explicitier selon une approche réaliste le sens des observations neuro-radiologique sous un format formel pour faciliter la réutilisation des données dans d'autres contextes. Nous avons réussi à mettre en place une preuve de concept pour structurer la sémantique de données d'observation en neuro-imagerie. Les résultats de notre expérimentation ont montré que l'annotation en format RDF de ces données nous permet d'extraire plus de connaissances à partir des informations textuelles qui sont disponibles dans la base de données REMBRANDT.

Dans ce travail, nous avons focalisé notre intérêt sur la modélisation des entités existantes en réalité et non pas sur la modélisation des artefacts humains. Nous avons fait ce choix pour explicitier la signification des critères d'imagerie tels que : les types des entités pathologiques étudiés, la localisation des entités pathologiques dans le cerveau, etc. Ainsi, nous avons proposé une catégorisation des différentes parties d'une tumeur cérébrale et des qualités ou dispositions qui sont mises en jeu dans l'évaluation radiologique. Nous avons utilisé des relations spatiales pour décrire des propriétés méreologiques comme "partie de" ainsi que des ontologies spécialisées comme FMA. La localisation des différentes parties pathologiques et anatomiques peut être très utile pour, par exemple, préparer à une opération chirurgicale, qui nécessite l'évaluation de localisation de la tumeur à traiter par rapport à certaines zones cérébrales fonctionnelles. Les entités et relations que nous proposons dans notre modèle peuvent être complétées par d'autres types d'informations et elles ne sont pas exhaustives pour couvrir le domaine de la neuro-imagerie. Dans ce travail,

nous avons représenté quelques aspects importants et communs dans le processus de l'observation en neuro-radiologie, à titre d'exemples, la région de la tumeur qui prend le contraste, la région de la tumeur qui ne prend pas de contraste, la partie kystique de la tumeur, etc.

Le module de génération des instances RDF permet d'associer les entités et relations relatives à chaque cas clinique. L'exploitation de cette base de données sémantique permet par exemple de catégoriser les tumeurs cérébrales selon les différentes descriptions attribuées et ainsi faire des recherches pour retrouver des informations sur des cas particuliers. Ces informations n'étant pas fournies dans la base REMBRANDT nous n'avons pas pu créer des instances relatives aux mesures du volume pour montrer l'intérêt du référencement explicite de ces mesures et de parties pathologiques qui leur sont associées. En résumé, ce travail d'expérimentation nous a permis de démontrer, au travers de quelques exemples d'exploitation, la valeur ajoutée de notre ontologie pour exprimer des informations qui n'étaient pas explicitement représentées.

Une première limitation de notre travail est qu'il ne prend pas en compte l'aspect de l'évolution des différentes entités référencées puisqu'elles peuvent changer de valeur (par exemple, le volume cérébral chez les patients atteints de la maladie d'Alzheimer peut changer au cours du temps), de nature (une tumeur bénigne peut devenir maligne) ou bien de nouvelles connaissances peuvent leur être associées. Ce point met en lumière le problème de la modélisation spatio-temporelle des entités avec l'ontologie BFO qui représente à lui seul un axe de recherche complexe [Adlassnig 2006]. Nous pensons que la poursuite de notre travail vers cette direction est une piste intéressante. Une deuxième limite de notre travail est qu'il n'applique pas encore un test de cohérence pour trouver "un consensus" entre les différentes annotations faites par les radiologues. Nous avons répartis les données de chaque radiologue dans des bases de données séparées, et nous avons accordé un seul ID à l'entité référencée ainsi la tumeur cérébrale de Mr.X par exemple peut être référencée de la façon suivante `_ct_900_00_1961`. Nous pensons que la meilleure façon est de réaliser un test de cohérence avant l'insertion dans la base de connaissance. Ainsi, nous pouvons garantir que la base soit cohérente et représentera la "réalité" (dans ce cas, la réalité est ce qui est vrai pour la plupart des radiologues). On doit noter que nous ne pouvons pas supprimer toute la subjectivité associée aux données d'observation étant donné qu'elles dépendent de plusieurs facteurs : les appareils et les protocoles de mesurage, les traitements d'imagerie, les observateurs, etc.

Nous avons remarqué que le temps de classification des instances est long (avec l'outil protégé) ce qui va compromettre sans doute l'efficacité des requêtes avec cet outil. Nous conseillons d'utiliser CORESE pour effectuer des tâches de raisonnement sur le corpus de données issu de la base REMBRANDT.



## 6.4 Évaluation du travail par rapport à l'existant

Ce travail, dont les motivations concernent à la fois l'utilisation des données d'imagerie pour la clinique (aide à la décision diagnostique, thérapeutique, aide au pronostic etc. dans le cadre de la médecine personnalisée) et pour la recherche (notamment découverte, validation et qualification de nouveaux biomarqueurs d'imagerie). Notre travail se situe dans la continuité de travaux antérieurs. Il s'inscrit dans une démarche visant à enrichir les standards et modèles d'information existants (DICOM, AIM, etc.) avec de nouveaux modes d'implémentation des méta-données d'imagerie (DICOM SR, DICOM header, données XML/AIM, etc.) afin de supporter le développement des futures biobanques d'images.

Notre modèle ontologique IBO est une initiative qui reste modeste par rapport aux spécifications des méta-données d'imagerie du standard DICOM. Cependant, ce que nous avons abordé dans cette thèse propose un socle ontologique générique qui peut être réutilisé et révisé pour fonder sur une base sémantique les concepts relatifs aux biomarqueurs d'imagerie qui sont inclus dans les comptes rendus DICOM SR. Cette orientation de description sémantique des biomarqueurs d'imagerie pour faciliter le partage des données entre des centres cliniques et des centres de recherche a été souligné dans plusieurs travaux [ESR 2015] [Bonmatí 2012], [Abramson 2015] [Levy 2012].

Par rapport au modèle d'information AIM, l'ontologie IBO et VASARI ont explicité d'une manière plus formelle et approfondie les concepts de ce modèle :

- L'ontologie IBO a couvert le "scope" des concepts provenant d'AIM et qui sont contenus dans le groupe de calcul et le groupe de "Mark up". IBO traite les observations qualitatives comme des biomarqueurs d'imagerie contrairement à AIM qui lui les considère comme des observations d'imagerie de type textuel. De plus, IBO ajoute plusieurs notions qui sont absentes dans AIM telles que la réalité biologique mesurée (le modèle ne prend en compte la réalité biologique de type physique), les protocoles appliqués pour exécuter les différents calculs de mesure, etc. Nous avons reformulé d'une façon ontologique toute la classification des formes graphiques du groupe "Mark up".
- L'ontologie VASARI a adopté l'approche réaliste pour représenter les concepts du groupe sémantique de l'image ; ainsi ce qui est implémenté dans notre ontologie peut être réutilisé pour enrichir les résultats d'imagerie (principalement l'absence d'entités physiques ou non physiques) décrits en AIM dans le contexte de la neuro-imagerie.

Notre travail IBO peut être réutilisé pour enrichir des travaux existants comme (1) le modèle MCI (Model for Clinical Information) du projet MEDICO en affichant des d'informations complémentaire aux résultats de classification des observations radiologiques ; (2) l'ontologie QIBO en formalisant et étendant ses concepts et relations en suivant notre méthodologie de construction ou bien en faisant des simples correspondances pour aligner les deux travaux sur une même vision, et l'ontologie BiomRKRS qui peut reprendre les concepts que nous avons retenus d'IBO pour représenter le processus de mesurage des biomarqueurs selon l'approche fondamentale

de BFO ; (3) le projet ANR NeuroLOG, qui a une vision plus large que nous, en exploitant les concepts que nous proposons dans notre travail pour la description des biomarqueurs. Nous soulignons que nous avons fait le choix de ne pas intégrer les deux ontologies développées dans le cadre de cette thèse pour garder l'aspect générique de l'ontologie IBO. Mais, les deux travaux sont intégrables facilement et l'ontologie de description des données d'observations peut compléter la description donnée par IBO pour les biomarqueurs d'imagerie neuro-radiologique.

Les bases de données de type RDF ont la particularité d'être flexibles, les données qu'elles contiennent peuvent changer de structure afin de répondre à des nouveaux besoins, ce qui n'est pas le cas pour les données du PACS. De plus, les jeux de données en RDF peuvent être mieux explorés via les points d'accès SPARQL. Ainsi, les résultats d'imagerie contenus dans des eCRF ou EHR peuvent être efficacement traités.

## Chapitre 7

# Conclusion générale et perspectives

La médecine personnalisée est fondamentalement basée sur l'association entre le profil biologique de chaque patient et les caractéristiques moléculaires de sa maladie. En imagerie médicale, des techniques d'imagerie avancées sont entrain d'émerger pour réaliser des analyses à l'échelle macroscopique afin d'améliorer la qualité des applications cliniques (diagnostiques, pronostiques, thérapeutiques) proposées aux patients. Pour cela, la sauvegarde informatisée à long terme des données brutes et dérivées d'imagerie, et des données génomiques qui sont issues de la recherche médicale et de la pratique clinique est devenue indispensable pour faciliter le développement et l'évaluation des biomarqueurs d'imagerie. Les systèmes d'archivage et de gestion des données d'aujourd'hui se trouvent face à deux défis majeurs : (1) être capable d'offrir des outils de partage et d'analyse de gros volumes de données sur de larges populations de patients afin d'explicitier/détecter les corrélations entre certains phénotypes d'imagerie et génotypes (2) être capable de formaliser les modèles de décision clinique pour pouvoir proposer des thérapies individualisées pour chaque patient durant sa prise en charge.

Le travail scientifique réalisé dans cette thèse aide particulièrement à enrichir sémantiquement les méta-données associées aux biomarqueurs d'imagerie pour répondre à certains enjeux d'analyse et de gestion de grandes masses de données comme à titre d'exemple :

- Faciliter la recherche sur les biomarqueurs d'imagerie dans une problématique clinique donnée ;
- Expliciter la signification biologique des biomarqueurs d'imagerie en explicitant le lien entre les aspects phénotypiques et génétiques spécifiques à la maladie étudiée ;
- Réaliser des études transversales autour des biomarqueurs d'imagerie ;
- Découvrir des nouveaux biomarqueurs d'imagerie en détectant certaines similitudes entre diverses maladies ainsi que les modèles animaux de ces maladies ;
- Intégration des biomarqueurs d'imagerie dans les "workflows" cliniques afin d'exécuter les étapes d'imagerie (de l'acquisition des examens d'imagerie à

la rédaction du compte rendu radiologique) nécessaires pour répondre aux mieux à la question clinique du médecin.

Notre travail a permis de mettre en lumière quelques défis soulevés dans le domaine de la standardisation des biomarqueurs d'imagerie tels que la définition d'une terminologie pour l'expression explicite de ce qu'ils représentent et la spécification des processus d'acquisition et de traitement pour garantir une meilleure compréhension des données de provenance des résultats. Le plus important pour nous n'est pas seulement de rendre accessible ces résultats de recherche et données cliniques à une large communauté scientifique mais aussi de faciliter la découverte de nouvelles connaissances. Par exemple, la fédération des données d'observation en utilisant le même schéma ontologique peut faciliter l'interprétation des différentes observations d'imagerie en fonction des propriétés physiques qu'elles étudient et des traitements d'imagerie qu'elles ont subis.

Les deux ontologies que nous avons développées méritent d'être utilisées dans la construction des entrepôts de données d'imagerie des patients issus de cas d'application réels ou dans la représentation sémantique des critères d'éligibilité dans les essais cliniques (voir [Weng 2010]). De plus, elles peuvent être d'une part compléter par l'inclusion des règles SWRL (Semantic Web Rule Language) du web sémantique et d'autre part, complétées par des modules de raisonnement interactifs pour automatiser leur gestion. Elles peuvent être aussi enrichies par des représentations de données incertaines pour renforcer leur capacité de raisonnement et mieux modéliser la réalité. Plusieurs travaux [John 2005] [Zheng 2008] ont été réalisés dans ce sens, certains ont appliqué les réseaux bayésiens aux modèles ontologiques pour implémenter des recommandations cliniques par exemple en cancérologie [Kahn 1997], en sclérose en plaques [Rodriguez 2012].

# Bibliographie

- [Abramson 2014] R.G. Abramson and T.E. Yankeelov. *Imaging biomarkers and surrogate endpoints in oncology clinical trials*. In *Functional Imaging in Oncology*, pages 29–42. Springer, 2014. (Cited on pages 31, 37 and 39.)
- [Abramson 2015] R.G. Abramson, K.R. Burton, J.Y.U. John-Paul, E.M. Scalzetti, T.E. Yankeelov, A.B. Rosenkrantz, M. Mendiratta-Lala, B.J. Bartholmai, D. Ganeshan, L. Lenchik *et al.* *Methods and challenges in quantitative imaging biomarker development*. *Academic radiology*, vol. 22, no. 1, pages 25–32, 2015. (Cited on pages 5, 26 and 153.)
- [Adlassnig 2006] K.P. Adlassnig, C. Combi, A.K. Das, E.T. Keravnou and G. Pozzi. *Temporal representation and reasoning in medicine : research directions and challenges*. *Artificial intelligence in medicine*, vol. 38, no. 2, pages 101–113, 2006. (Cited on page 152.)
- [Ah-See 2008] M.L.W. Ah-See, A. Makris, N.J. Taylor, M. Harrison, P.I. Richman, R.J. Burcombe, J.J. Stirling, J.A. d’Arcy, D.J. Collins, M.R. Pittam *et al.* *Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer*. *Clinical Cancer Research*, vol. 14, no. 20, pages 6580–6589, 2008. (Cited on page 27.)
- [Amdouni 2015] E. Amdouni, Y. Morvan and B. Gibaud. *Towards an Imaging Biomarker Ontology Based on the Open Biological and Biomedical Ontologies Foundry*. In *SWA4TLS : Semantic Web Applications Tools for Life Sciences*, pages 193–194, 2015. (Cited on page 14.)
- [Amdouni 2016a] E. Amdouni and B. Gibaud. *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, 2016. (Cited on page 14.)
- [Amdouni 2016b] E. Amdouni and B. Gibaud. *Conception d’une ontologie générique pour la représentation sémantique des biomarqueurs d’imagerie*. In *JFO : Journées Francophones sur les ontologies*, pages 12–24, 2016. (Cited on page 14.)
- [Archive 2013] The Cancer Imaging Archive. *Wiki for the VASARI feature set. VASARI Research Project*, 2013. (Cited on page 36.)

- [Arp 2015] R. Arp, B. Smith and A.D. Spear. Building ontologies with basic formal ontology. Mit Press, 2015. (Cited on page 97.)
- [Asslaber 2007] M. Asslaber and K. Zatloukal. *Biobanks : transnational, European and global networks*. Briefings in functional genomics & proteomics, vol. 6, no. 3, pages 193–201, 2007. (Cited on page 48.)
- [Ball 2010] J.R. Ball, C.M. Michelet *et al.* Evaluation of biomarkers and surrogate endpoints in chronic disease. National Academies Press, 2010. (Cited on pages 41, 43 and 92.)
- [Bandrowski 2016] A. Bandrowski, R. Brinkman, M. Brochhausen, M.H. Brush, B. Bug, M.C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier *et al.* *The ontology for biomedical investigations*. PloS one, vol. 11, no. 4, page e0154556, 2016. (Cited on pages 7, 99, 100 and 102.)
- [Barajas J.R. 2009] R.F. Barajas J.R., J.S. Chang, M.R. Segal, A.T. Parsa, M.W. McDermott, M.S. Berger and S. Cha. *Differentiation of recurrent glioblastoma multiforme from radiation necrosis after external beam radiation therapy with dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging*. Radiology, vol. 253, no. 2, pages 486–496, 2009. (Cited on page 36.)
- [Bard 2005] J. Bard, S.Y. Rhee and M. Ashburner. *An ontology for cell types*. Genome biology, vol. 6, no. 2, page R21, 2005. (Cited on page 76.)
- [Basser 2011] P.J. Basser and C. Pierpaoli. *Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI*. Journal of magnetic resonance, vol. 213, no. 2, pages 560–570, 2011. (Cited on page 26.)
- [Bennett 2013] B. Bennett, V. Chaudhri and N. Dinesh. *A vocabulary of topological and containment relations for a practical biological ontology*. In International Conference on Spatial Information Theory, pages 418–437. Springer, 2013. (Cited on page 128.)
- [Bisdas 2009] S. Bisdas, M. Kirkpatrick, P. Giglio, C. Welsh, M.V. Spampinato and Z. Rumboldt. *Cerebral blood volume measurements by perfusion-weighted MR imaging in gliomas : ready for prime time in predicting short-term outcome and recurrent disease ?* American Journal of Neuroradiology, vol. 30, no. 4, pages 681–688, 2009. (Cited on page 36.)
- [Bittner 2003] T. Bittner and B. Smith. *Formal ontologies for space and time*. IFO-MIS, Department of Philosophy. Leipzig, Buffalo, University of Leipzig, University at Buffalo and NCGIA, vol. 17, 2003. (Cited on page 94.)
- [Bittner 2004] T. Bittner and B. Smith. *Normalizing medical ontologies using basic formal ontology*. 2004. (Cited on page 94.)
- [Bjartmar 2000] C. Bjartmar, G. Kidd, S. Mörk, R. Rudick and B.D. Trapp. *Neurological disability correlates with spinal cord axonal loss and reduced N-acetyl aspartate in chronic multiple sclerosis patients*. Annals of neurology, vol. 48, no. 6, pages 893–901, 2000. (Cited on page 24.)

- [Bomers 2014] J.G.R. Bomers and J.O. Barentsz. *Standardization of multiparametric prostate MR imaging using PI-RADS*. BioMed research international, vol. 2014, 2014. (Cited on page 36.)
- [Bonmatí 2012] L.M. Bonmatí, A. Alberich-Bayarri, G. García-Martí, R.S. Requena, C.P. Castillo, J.M.C. Sierra and J.V.M. Herrera. *Imaging biomarkers, quantitative imaging, and bioengineering*. Radiología (English Edition), vol. 54, no. 3, pages 269–278, 2012. (Cited on pages 41 and 153.)
- [Boone 2007] J.M. Boone. *Radiological interpretation 2020 : toward quantitative image assessment*. Medical physics, vol. 34, no. 11, pages 4173–4179, 2007. (Cited on page 32.)
- [Bosmans 2012] J.M.L. Bosmans, L. Peremans, M. Menni, A.M. De Schepper, P.O. Duyck and P.M. Parizel. *Structured reporting : if, why, when, how and at what expense ? Results of a focus group meeting of radiology professionals from eight countries*. Insights into imaging, vol. 3, no. 3, pages 295–302, 2012. (Cited on page 66.)
- [Bradshaw 2013] S. Bradshaw, D. Brickley, L.J.G. Castro, T. Clark, T. Cole, A. Dessen P.and Gerber, A. Isaac, J. Jett, T. Habinget *al. Open Annotation Data Model : community draft*. 2013. (Cited on page 76.)
- [Brinkman 2010] R.R. Brinkman, M. Courtot, D. Derom, J. Fostel, Y. He, P.W. Lord, J. Malone, H.E. Parkinson, B. Peters, P. Rocca-Serraet *al. Modeling biomedical experimental processes with OBI*. J. Biomedical Semantics, vol. 1, no. S-1, page S7, 2010. (Cited on page 101.)
- [Brodbelt 2014] A. Brodbelt. *Clinical applications of imaging biomarkers. Part 2. The neurosurgeon's perspective*. The British journal of radiology, 2014. (Cited on page 32.)
- [Brunnbauer 2013] M. Brunnbauer. *DICOM metadata as RDF*. In GI-Jahrestagung, pages 1796–1804, 2013. (Cited on page 66.)
- [Bruns 2009] A. Bruns, B. Künnecke, C. Risterucci, J.L. Moreau and M. von Kienlin. *Validation of cerebral blood perfusion imaging as a modality for quantitative pharmacological MRI in rats*. Magnetic resonance in medicine, vol. 61, no. 6, pages 1451–1458, 2009. (Cited on page 43.)
- [Buckler 2011a] A.J. Buckler and R. Boellaard. *Standardization of quantitative imaging : the time is right, and 18F-FDG PET/CT is a good place to start*. Journal of Nuclear Medicine, vol. 52, no. 2, pages 171–172, 2011. (Cited on pages 39 and 43.)
- [Buckler 2011b] A.J. Buckler, L. Bresolin, N.R. Dunnick and D.C. Sullivan. *A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging*. Radiology, vol. 258, no. 3, pages 906–914, 2011. (Cited on pages 5, 20, 21 and 32.)
- [Buckler 2011c] A.J. Buckler, L. Bresolin, N.R. Dunnick and D.C. Sullivan. *Quantitative imaging test approval and biomarker qualification : interrelated but*

- distinct activities*. Radiology, vol. 259, no. 3, pages 875–884, 2011. (Cited on page 45.)
- [Buckler 2013a] A.J. Buckler, M. Ouellette, J. Danagoulian, G. Wernsing, T.T. Liu, E. Savig, B.E. Suzek and Paik D. Rubin D.L. *Quantitative imaging biomarker ontology (QIBO) for knowledge representation of biomedical imaging biomarkers*. Journal of digital imaging, vol. 26, no. 4, pages 630–641, 2013. (Cited on pages 7, 79 and 82.)
- [Buckler 2013b] A.J. Buckler, D. Paik, M. Ouellette, J. Danagoulian, G. Wernsing and B.E. Suzek. *A novel knowledge representation framework for the statistical validation of quantitative imaging biomarkers*. Journal of digital imaging, vol. 26, no. 4, pages 614–629, 2013. (Cited on pages 20 and 47.)
- [Buvat 2007] I. Buvat. *Understanding the limitations of SUV*. Medecine Nucleaire, vol. 31, pages 165–172, 2007. (Cited on page 33.)
- [Ceusters 2005] W. Ceusters and B. Smith. *Tracking referents in electronic health records*. Studies in health technology and informatics, vol. 116, page 71, 2005. (Cited on page 126.)
- [Ceusters 2006] W. Ceusters, P. Elkin and B. Smith. *Referent tracking : The problem of negative findings*. Studies in health technology and informatics, vol. 124, page 741, 2006. (Cited on pages 126, 128 and 142.)
- [Ceusters 2012] W. Ceusters. *An information artifact ontology perspective on data collections and associated representational artifacts*. In MIE, pages 68–72, 2012. (Cited on page 102.)
- [Ceusters 2015] W. Ceusters and B. Smith. *Biomarkers in the Ontology for General Medical Science*. In Ronald Cornet (ed.), Digital Healthcare Empowering Europeans, vol. 512, no. 8, pages 155–159, 2015. (Cited on pages 7, 84, 85, 92 and 122.)
- [Chan 2007] M. Chan, S. Thellier, B. Heuga, J. Hannequin, A. Pluvinage, J.M. Tourani, F. Fruge, B. Rouillet, R. Vialle and J.P. Tasu. *ONCO-WP-8 Comment interpreter les bilans de surveillance d'un cancer en imagerie ?* Journal de Radiologie, vol. 88, no. 10, page 1561, 2007. (Cited on pages 5 and 34.)
- [Channin 2010] D.S. Channin, P. Mongkolwat, V. Kleper, K. Sepukar and D.L. Rubin. *The caBIG annotation and image markup project*. Journal of digital imaging, vol. 23, no. 2, pages 217–225, 2010. (Cited on pages 6, 49, 67, 68 and 69.)
- [Choong 2012] M.K. Choong and G. Tsafnat. *The implications of biomarker evidence for systematic reviews*. BMC medical research methodology, vol. 12, no. 1, page 176, 2012. (Cited on page 32.)
- [Ciccarese 2011] P. Ciccarese, M. Ocana, L.J.G. Castro and Clark T. Das S. *An open annotation ontology for science on web 3.0*. Journal of biomedical semantics, vol. 2, no. 2, page S4, 2011. (Cited on page 76.)



- [Ciccarese 2012] P. Ciccarese, M. Ocana and T. Clark. *Open semantic annotation of scientific publications using DOME0*. Journal of biomedical semantics, vol. 3, no. 1, page S1, 2012. (Cited on page 78.)
- [Ciccarese 2014] P. Ciccarese. *integrating open annotation with any domain ontology*. 2014. (Cited on pages 7 and 78.)
- [Cimino 2006] J.J. Cimino. *In defense of the Desiderata*. Journal of biomedical informatics, vol. 39, no. 3, pages 299–306, 2006. (Cited on pages 124 and 142.)
- [Clunie 2000] D.A. Clunie. DICOM structured reporting. PixelMed Publishing, 2000. (Cited on page 60.)
- [Clunie 2007a] D.A. Clunie. *DICOM structured reporting and cancer clinical trials results*. Cancer informatics, vol. 4, page 33, 2007. (Cited on pages 6, 49 and 65.)
- [Clunie 2007b] D.A. Clunie. *Encoding of Oncology Clinical Trial Read Results in DICOM SR*. [http://qibawiki.rsna.org/images/f/fc/ClinicalTrialResults\\_20111123\\_QIBA.doc](http://qibawiki.rsna.org/images/f/fc/ClinicalTrialResults_20111123_QIBA.doc), [Online] accessed 2011-11-23, 2007. (Cited on pages 64 and 105.)
- [Clunie 2012] D.A. Clunie. *DICOM Correction Proposal*. [ftp://medical.nema.org/medical/dicom/final/cp1112\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/cp1112_ft.pdf), 2012. (Cited on pages 7 and 96.)
- [Consortium 2008] Gene Ontology Consortium *et al.* *The gene ontology project in 2008*. Nucleic acids research, vol. 36, no. suppl 1, pages D440–D444, 2008. (Cited on pages 75, 100 and 111.)
- [Craddock 2013] R.C. Craddock, S. Jbabdi, C.G. Yan, Vogelstein J.T., F.X. Castellanos, M.A. Di, C. Kelly, K. Heberlein, S. Colcombe and M.P. Milham. *Imaging human connectomes at the macroscale*. Nature methods, vol. 10, no. 6, pages 524–539, 2013. (Cited on page 27.)
- [Crommelin 2011] D.J.A. Crommelin, G. Storm and P. Luijten. *‘Personalised medicine’ through ‘personalised medicines’ : time to integrate advanced, non-invasive imaging approaches and smart drug delivery systems*. International journal of pharmaceutics, vol. 415, no. 1, pages 5–8, 2011. (Cited on page 19.)
- [Degtyarenko 2008] K. Degtyarenko, P. De Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M.l Guedj and M. Ashburner. *ChEBI : a database and ontology for chemical entities of biological interest*. Nucleic acids research, vol. 36, no. suppl 1, pages D344–D350, 2008. (Cited on page 100.)
- [Desar 2009] I.M.E. Desar, C.M.L Van Herpen, H.W.M Van Laarhoven, J.O. Barentsz, W.J.G. Oyen and W.T.A. Van Der Graaf. *Beyond RECIST : molecular and functional imaging techniques for evaluation of response to targeted therapy*. Cancer treatment reviews, vol. 35, no. 4, pages 309–321, 2009. (Cited on page 32.)

- [Després 2007] S. Després and S. Mhiri. *Ontology Visualization Tool for Indexing DICOM Structured Reporting (SR) Documents*. In 10th International Protégé Conference, pages 150–153, 2007. (Cited on pages 6, 66 and 67.)
- [Di Chiro 1988a] G. Di Chiro and R.A. Brooks. *PET-FDG of untreated and treated cerebral gliomas*. *Journal of Nuclear Medicine*, vol. 29, no. 3, pages 421–422, 1988. (Cited on page 30.)
- [Di Chiro 1988b] G. Di Chiro, E. Oldfield, D.C. Wright, D. De Michele, D.A. Katz, N.J. Patronas, J.L.S.M. Doppman and Kufta C.V. Ito M. *Cerebral necrosis after radiotherapy and/or intraarterial chemotherapy for brain tumors : PET and neuropathologic studies*. *American Journal of Roentgenology*, vol. 150, no. 1, pages 189–197, 1988. (Cited on page 30.)
- [Dickerson 2005] B.C. Dickerson and R.A. Sperling. *Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer’s disease*. *NeuroRx*, vol. 2, no. 2, pages 348–360, 2005. (Cited on page 32.)
- [Diehn 2008] M. Diehn, C. Nardini, D.S. Wang, S. McGovern, M. Jayaraman, Y.U. Liang, K. Aldape, S. Cha and M.D. Kuo. *Identification of noninvasive imaging surrogates for brain tumor gene-expression modules*. *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pages 5213–5218, 2008. (Cited on page 36.)
- [DiSantis 1986] D.J. DiSantis. *Early American radiology : the pioneer years*. *American Journal of Roentgenology*, vol. 147, no. 4, pages 850–853, 1986. (Cited on page 47.)
- [Downing 2001] G. Downing. *Biomarkers Definitions Working Group. Biomarkers and Surrogate Endpoints*. *Clinical Pharmacology & Therapeutics*, vol. 69, pages 89–95, 2001. (Cited on pages 6, 31 and 45.)
- [Eisenhauer 2009] E.A. Eisenhauer, Bogaerts J. Therasse P., L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, Mooney M.*et al.* *New response evaluation criteria in solid tumours : revised RECIST guideline (version 1.1)*. *European journal of cancer*, vol. 45, no. 2, pages 228–247, 2009. (Cited on pages 23, 32 and 33.)
- [ESR 2010] ESR. *White paper on imaging biomarkers*. *Insights into imaging*, vol. 1, no. 2, pages 42–45, 2010. (Cited on pages 5, 20, 30, 31, 32 and 92.)
- [ESR 2011] ESR. *Medical imaging in personalised medicine : a white paper of the research committee of the European Society of Radiology (ESR)*. *Insights into imaging*, vol. 2, no. 6, pages 621–630, 2011. (Cited on pages 18, 45 and 48.)
- [ESR 2013] ESR. *ESR statement on the stepwise development of imaging biomarkers*. *Insights into imaging*, vol. 4, no. 2, pages 147–152, 2013. (Cited on pages 6, 41 and 44.)
- [ESR 2015] ESR. *ESR position paper on imaging biobanks*. *Insights into imaging*, vol. 6, no. 4, pages 403–410, 2015. (Cited on pages 48, 70 and 153.)

- [FDA 2005] FDA. *Biomarker qualification program*. <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentTool/QualificationProgram/BiomarkerQualificationProgram/ucm535922.html>, 2005. (Cited on pages 6 and 38.)
- [Filippi 2014] M. Filippi, P.a Preziosa and M.A. Rocca. *Magnetic resonance outcome measures in multiple sclerosis trials : time to rethink ?* Current opinion in neurology, vol. 27, no. 3, pages 290–299, 2014. (Cited on page 32.)
- [Filippi 2016] M. Filippi, M.A. Rocca, O. Ciccarelli, N. De Stefano, N. Evangelou, L. Kappos, A. Rovira, Tintorè M. Sastre-Garriga J., J.L. Frederiksen *et al.* *MRI criteria for the diagnosis of multiple sclerosis : MAGNIMS consensus guidelines*. The Lancet Neurology, vol. 15, no. 3, pages 292–303, 2016. (Cited on pages 6 and 53.)
- [Fletcher 2008] J.W. Fletcher, B. Djulbegovic, H.P. Soares, B.A. Siegel, V.J. Lowe, G.H. Lyman, R.E. Coleman, R. Wahl, J.C Paschold, N. Avril *et al.* *Recommendations on the use of 18F-FDG PET in oncology*. Journal of Nuclear Medicine, vol. 49, no. 3, pages 480–508, 2008. (Cited on page 34.)
- [Forward 2012] Look ESF Forward. *Personalised medicine for the European citizen*. Towards more precise medicine for the diagnosis, treatment and prevention of disease (iPM). European Science Foundation, Strasbourg November, 2012. (Cited on page 18.)
- [Frank 2003] R. Frank and R. Hargreaves. *Clinical biomarkers in drug discovery and development*. Nature reviews. Drug discovery, vol. 2, no. 7, page 566, 2003. (Cited on page 37.)
- [Garcia-Donas 2013] J. Garcia-Donas, C. Rodriguez-Antona and E. Jonasch. *Molecular markers to predict response to therapy*. In Seminars in oncology, volume 40, pages 444–458. Elsevier, 2013. (Cited on pages 31 and 32.)
- [Garrett 1999] M.D. Garrett and P. Workman. *Discovering novel chemotherapeutic drugs for the third millennium*. European Journal of Cancer, vol. 35, no. 14, pages 2010–2030, 1999. (Cited on page 37.)
- [Gibaud 2011] B. Gibaud, G. Kassel, M. Dojat, B. Batrancourt, F. Michel, A. Gaignard and J. Montagnat. *NeuroLOG : sharing neuroimaging data using an ontology-based federated approach*. In AMIA... Annual Symposium proceedings [electronic resource]/AMIA Symposium. AMIA Symposium., volume 2011, pages 472–80, 2011. (Cited on page 78.)
- [Giorgio 2013] A. Giorgio and N. De Stefano. *Clinical use of brain volumetry*. Journal of Magnetic Resonance Imaging, vol. 37, no. 1, pages 1–14, 2013. (Cited on page 24.)
- [Gkoutos 2012] G.V. Gkoutos, P.N. Schofield and R. Hoehndorf. *The Units Ontology : a tool for integrating units of measurement in science*. Database, vol. 2012, page bas033, 2012. (Cited on page 103.)

- [Green 2008] E.M. Green, G. Yothers and D.J. Sargent. *Surrogate endpoint validation : statistical elegance versus clinical relevance*. Statistical methods in medical research, 2008. (Cited on page 44.)
- [Grenon 2004] P. Grenon, B. Smith and L. Goldberg. *Biodynamic ontology : applying BFO in the biomedical domain*. Studies in health technology and informatics, pages 20–38, 2004. (Cited on page 94.)
- [Gruber 1993] T.R. Gruber. *A translation approach to portable ontology specifications*. Knowledge acquisition, vol. 5, no. 2, pages 199–220, 1993. (Cited on page 71.)
- [Guarino 1998] N. Guarino *et al.* *Formal ontology and information systems*. In Proceedings of FOIS, volume 98, pages 81–97, 1998. (Cited on page 71.)
- [Gutman 2013] D.A. Gutman, L.A.D. Cooper, S.N. Hwang, C.A. Holder, J. Gao, T.D. Aurora, W.D. Dunn Jr, L. Scarpace, T. Mikkelsen, R. Jain *et al.* *MR imaging predictors of molecular profile and survival : multi-institutional study of the TCGA glioblastoma data set*. Radiology, vol. 267, no. 2, pages 560–569, 2013. (Cited on pages 8, 36 and 127.)
- [Haacke 1999] E.M. Haacke, R.W. Brown, R. Thompson, R. Venkatesan *et al.* *Magnetic resonance imaging : physical principles and sequence design*, volume 82. Wiley-Liss New York :, 1999. (Cited on page 24.)
- [Hanahan 2000] D. Hanahan and R.A. Weinberg. *The hallmarks of cancer*. cell, vol. 100, no. 1, pages 57–70, 2000. (Cited on page 33.)
- [Harris 1986] R.K. Harris. *Nuclear magnetic resonance spectroscopy*. 1986. (Cited on page 27.)
- [Hewitt 2013] R. Hewitt and P. Watson. *Defining biobank*. Biopreservation and biobanking, vol. 11, no. 5, pages 309–315, 2013. (Cited on page 48.)
- [Hu 2012] L.S. Hu, J.M. Eschbacher, A.C. Dueck, J.E. Heiserman, S. Liu, J.P. Karis, K.A. Smith, W.R. Shapiro, D.S. Pinnaduwage, S.W. Coonset *et al.* *Correlations between perfusion MR imaging cerebral blood volume, microvessel quantification, and clinical outcome using stereotactic analysis in recurrent high-grade glioma*. American Journal of Neuroradiology, vol. 33, no. 1, pages 69–76, 2012. (Cited on page 36.)
- [Huettel 2004] S.A. Huettel, A.W. Song and G. McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004. (Cited on page 28.)
- [Hussein 2004] R. Hussein, U. Engelmann, A. Schroeter and H.P. Meinzer. *DICOM Structured Reporting : Part 2. Problems and Challenges in Implementation for PACS Workstations 1*. Radiographics, vol. 24, no. 3, pages 897–909, 2004. (Cited on page 60.)
- [Hutter 2003] A. Hutter, K.E. Schwetye, A.J. Bierhals and R.C. McKinstry. *Brain neoplasms : epidemiology, diagnosis, and prospects for cost-effective imaging*. Neuroimaging clinics of North America, vol. 13, no. 2, pages 237–250, 2003. (Cited on page 32.)

- [Huynh 2014] T.J. Huynh, M.L. Flaherty, D.J. Gladstone, J.P. Broderick, A.M. Demchuk, D. Dowlatshahi, A. Meretoja, S.M. Davis, P.J. Mitchell, G.A. Tomlinson *et al.* *Multicenter accuracy and interobserver agreement of spot sign identification in acute intracerebral hemorrhage*. *Stroke*, vol. 45, no. 1, pages 107–112, 2014. (Cited on page 43.)
- [Jaffer 2005] F.A. Jaffer and R. Weissleder. *Molecular imaging in the clinical arena*. *Jama*, vol. 293, no. 7, pages 855–862, 2005. (Cited on page 20.)
- [Jain 2014] R. Jain, L.M. Poisson, D. Gutman, L. Scarpace, S.N. Hwang, C.A. Holder, M. Wintermark, A. Rao, R.R. Colen, J. Kirby *et al.* *Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers : focus on the nonenhancing component of the tumor*. *Radiology*, vol. 272, no. 2, pages 484–493, 2014. (Cited on page 117.)
- [John 2005] R.I. John and P.R. Innocent. *Modeling uncertainty in clinical diagnosis using fuzzy logic*. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 6, pages 1340–1350, 2005. (Cited on page 156.)
- [Joint Committee for Guides in Metrology 2012] Joint Committee for Guides in Metrology. *International Vocabulary of Metrology – basic and general concepts and associated terms*. [www.nist.gov/pml/div688/grp40/upload/international-vocabularyofmetrology.pdf](http://www.nist.gov/pml/div688/grp40/upload/international-vocabularyofmetrology.pdf), [online] accessed 2015-06-22, 2012. (Cited on pages 6, 47 and 48.)
- [Kahn Jr 2009] C.E. Kahn Jr, C.P. Langlotz, E.S. Burnside, J.A. Carrino, D.S. Channin, D.M. Hovsepian and D.L. Rubin. *Toward best practices in radiology reporting 1*. *Radiology*, vol. 252, no. 3, pages 852–856, 2009. (Cited on page 61.)
- [Kahn 1997] C.E. Kahn, L.M. Roberts, K.A. Shaffer and P. Haddawy. *Construction of a Bayesian network for mammographic diagnosis of breast cancer*. *Computers in biology and medicine*, vol. 27, no. 1, pages 19–29, 1997. (Cited on page 156.)
- [Kao 2013] H.W. Kao, S.W. Chiang, H.W. Chung, F.Y. Tsai and C.Y. Chen. *Advanced MR imaging of gliomas : an update*. *BioMed research international*, vol. 2013, 2013. (Cited on pages 5, 6, 28, 32 and 35.)
- [Katz 2004] R. Katz. *Biomarkers and surrogate markers : an FDA perspective*. *NeuroRx*, vol. 1, no. 2, pages 189–195, 2004. (Cited on pages 31 and 44.)
- [Kazerooni 2016] E.A. Kazerooni, M.R. Armstrong, J.K. Amorosa, D. Hernandez, L.A. Liebscher, H. Nath, M.F. McNitt-Gray, E.J. Stern and P.A. Wilcox. *ACR CT accreditation program and the lung cancer screening program designation*. *Journal of the American College of Radiology*, vol. 13, no. 2, pages R30–R34, 2016. (Cited on page 36.)
- [Kessler 2015] L.G. Kessler, H.X. Barnhart, A.J. Buckler, K.R. Choudhury, M.V. Kondratovich, A. Toledano, A.R. Guimaraes, R. Filice, Z. Zhang and D.C. Sullivan. *The emerging science of quantitative imaging biomarkers termino-*

- logy and definitions for scientific studies and regulatory submissions*. Statistical methods in medical research, vol. 24, no. 1, pages 9–26, 2015. (Cited on pages 20, 31 and 46.)
- [Kevin 2016] O. Kevin, Daniel S., Eric P., Nancy O., Patricia C. Michael B., Cathy E., Brian G., Alexander G., Edward J., Lisa K. and Nicholas P. Paul K. Fe-roze M. *QIBA Process Committee : Overview and Status Update*. 2016. (Cited on pages 6 and 40.)
- [Kinahan 2010] P.E. Kinahan and J.W. Fletcher. *Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy*. In Seminars in Ultrasound, CT and MRI, volume 31, pages 496–505. Elsevier, 2010. (Cited on page 34.)
- [Kopans 1992] D.B. Kopans. *Standardized mammography reporting*. Radiologic Clinics of North America, vol. 30, no. 1, pages 257–264, 1992. (Cited on page 36.)
- [Korenblum 2011] D. Korenblum, D. Rubin, S. Napel, C. Rodriguez and C. Beaulieu. *Managing biomedical image metadata for search and retrieval of similar images*. Journal of digital imaging, vol. 24, no. 4, pages 739–748, 2011. (Cited on page 68.)
- [Kumar 2003] A. Kumar and B. Smith. *The universal medical language system and the gene ontology : Some critical reflections*. In Annual Conference on Artificial Intelligence, pages 135–148. Springer, 2003. (Cited on page 75.)
- [la Fougère 2011] C. la Fougère, B. Suchorska, P. Bartenstein, F.W. Kreth and J. Tonn. *Molecular imaging of gliomas with PET : opportunities and limitations*. Neuro-oncology, page nor054, 2011. (Cited on page 34.)
- [Lacson 2012] R. Lacson, K.P. Andriole, L.M. Prevedello and R. Khorasani. *Information from searching content with an ontology-utilizing toolkit (iSCOUT)*. Journal of digital imaging, vol. 25, no. 4, pages 512–519, 2012. (Cited on page 61.)
- [Larson 1999] S.M.E.Y. Larson, T. Akhurst, M. Mazumdar, H.A. Macapinlac, R.D. Finn, C. Casilla, M. Fazzari, N. Srivastava, H.W.D. Yeunget al. *Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging : the visual response score and the change in total lesion glycolysis*. Clinical Positron Imaging, vol. 2, no. 3, pages 159–171, 1999. (Cited on page 30.)
- [Lavin 1980] P.T. Lavin and G. Flowerdew. *Studies in variation associated with the measurement of solid tumors*. Cancer, vol. 46, no. 5, pages 1286–1290, 1980. (Cited on page 33.)
- [Le Bihan 1986] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis and M. Laval-Jeantet. *MR imaging of intravoxel incoherent motions : application to diffusion and perfusion in neurologic disorders*. Radiology, vol. 161, no. 2, pages 401–407, 1986. (Cited on page 26.)

- [Le Bihan 2001] D. Le Bihan, J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko and H. Chabriat. *Diffusion tensor imaging : concepts and applications*. Journal of magnetic resonance imaging, vol. 13, no. 4, pages 534–546, 2001. (Cited on page 26.)
- [Le Bihan 2003] D. Le Bihan. *Looking into the functional architecture of the brain with diffusion MRI*. Nature Reviews Neuroscience, vol. 4, no. 6, pages 469–480, 2003. (Cited on page 26.)
- [Le Bihan 2013] D. Le Bihan. *Apparent diffusion coefficient and beyond : what diffusion MR imaging can tell us about tissue structure*, 2013. (Cited on page 26.)
- [Levy 2012] M.A. Levy, J.B. Freymann, J.S. Kirby, A. Fedorov, F.M. Fennessy, S.A. Eschrich, A.E. Berglund, D.A. Fenstermacher, Guo X. Tan Y. *et al. Informatics methods to enable sharing of quantitative imaging research data*. Magnetic resonance imaging, vol. 30, no. 9, pages 1249–1256, 2012. (Cited on pages 11, 49, 50, 142 and 153.)
- [Lynda 2008] T. Lynda. *Ontologie de partage de données et d'outils de traitement dans le domaine de la neuroimagerie*. PhD thesis, Université de Rennes 1, 2008. (Cited on pages 7 and 80.)
- [Malone 2010] J. Malone, E. Holloway, T. Adamusiak, M. Kapushesky, J. Zheng, N. Kolesnikov, A. Zhukova, A. Brazma and H. Parkinson. *Modeling sample variables with an Experimental Factor Ontology*. Bioinformatics, vol. 26, no. 8, pages 1112–1118, 2010. (Cited on page 82.)
- [Marrer 2007] E. Marrer and F. Dieterle. *Promises of Biomarkers in Drug Development—A Reality Check*. Chem Biol Drug Des, vol. 69, pages 381–394, 2007. (Cited on page 45.)
- [Masolo 2003] C. Masolo, S. Borgo, A. Gangemi, N. Guarino and A. Oltramari. *Wonderweb deliverable d18, ontology library (final)*. ICT project, vol. 33052, 2003. (Cited on pages 78 and 110.)
- [McBride 2002] B. McBride. *Jena : A semantic web toolkit*. IEEE Internet computing, vol. 6, no. 6, pages 55–59, 2002. (Cited on page 129.)
- [Miller 2004] D.H. Miller. *Biomarkers and surrogate outcomes in neurodegenerative disease : lessons from multiple sclerosis*. NeuroRx, vol. 1, no. 2, pages 284–294, 2004. (Cited on page 32.)
- [Moertel 1976] C.G. Moertel and J.A. Hanley. *The effect of measuring error on the results of therapeutic trials in advanced cancer*. Cancer, vol. 38, no. 1, pages 388–394, 1976. (Cited on page 33.)
- [Mungall 2007] C. Mungall, G. Gkoutos, N. Washington and S. Lewis. *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, 2007. (Cited on pages 99 and 105.)
- [Mungall 2013] C.J. Mungall. *A critique of temporalized relations*. Tech. rep. Berkeley : Genomics Division, Lawrence Berkeley National Laboratory, pages 1–10, 2013. (Cited on page 98.)

- [Noumeir 2003] R. Noumeir. *DICOM Structured Report Document Type Definition*. Trans. Info. Tech. Biomed., vol. 7, no. 4, pages 318–328, December 2003. (Cited on page 60.)
- [Noy 2009] N.F. Noy, N.H. Shah, P.L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D.L. Rubin, M.A. Storey, C.G.G. Chuteet al. *BioPortal : ontologies and integrated data resources at the click of a mouse*. Nucleic acids research, vol. 37, no. suppl 2, pages W170–W173, 2009. (Cited on page 74.)
- [Oberkampff 2012] H. Oberkampff, S. Zillner, B. Bauer and M. Hammon. *Interpreting Patient Data using Medical Background Knowledge*. ICBO, vol. 897, page 3, 2012. (Cited on pages 7, 86 and 87.)
- [Oberkampff 2016] H. Oberkampff, S. Zillner, J.A. Overton, B. Bauer, A. Cavallaro, M. Uder and M. Hammon. *Semantic representation of reported measurements in radiology*. BMC medical informatics and decision making, vol. 16, no. 1, page 5, 2016. (Cited on page 86.)
- [O’Connor 2008] J.P.B. O’Connor, A. Jackson, M.C. Asselin, D.L. Buckley, G.J.M. Parker and G.C. Jayson. *Quantitative imaging biomarkers in the clinical development of targeted therapeutics : current and future perspectives*. The lancet oncology, vol. 9, no. 8, pages 766–776, 2008. (Cited on pages 31 and 37.)
- [Ofoghi 2014] B. Ofoghi, G.H.L. Campos, K. Verspoor and F.J.M. Sanchez. *BiomR-KRS : a biomarker retrieval and knowledge reasoning system*. In Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management-Volume 153, pages 31–39. Australian Computer Society, Inc., 2014. (Cited on pages 7, 81 and 83.)
- [Petersen 2013] S.E. Petersen, P.M. Matthews, F. Bamberg, D.A. Bluemke, J.M. Francis, M.G. Friedrich, P. Leeson, E. Nagel, S. Plein, F.E. Rademakers et al. *Imaging in population science : cardiovascular magnetic resonance in 100,000 participants of UK Biobank-rationale, challenges and approaches*. Journal of Cardiovascular Magnetic Resonance, vol. 15, no. 1, page 46, 2013. (Cited on page 48.)
- [Popescu 2013] V. Popescu, F. Agosta, H.E. Hulst, I.C. Sluimer, D.L. Knol, M.P. Sormani, C. Enzinger, S. Ropele, J. Alonso, J. Sastre-Garriga et al. *Brain atrophy and lesion load predict long term disability in multiple sclerosis*. Journal of Neurology, Neurosurgery & Psychiatry, vol. 84, no. 10, pages 1082–1091, 2013. (Cited on page 24.)
- [Porz 2014] N. Porz, S. Bauer, A. Pica, P. Schucht, J. Beck, R.K. Verma, J. Slotboom, M. Reyes and R. Wiest. *Multi-modal glioblastoma segmentation : man versus machine*. PloS one, vol. 9, no. 5, page e96873, 2014. (Cited on page 143.)
- [Prescott 2013] J.W. Prescott. *Quantitative imaging biomarkers : the application of advanced image processing and analysis to clinical and preclinical decision*



- making*. Journal of digital imaging, vol. 26, no. 1, pages 97–108, 2013. (Cited on page 31.)
- [Provenzale 2002] J.M. Provenzale, G.R. Wang, T. Brenner, J.R. Petrella and A.G. Sorensen. *Comparison of permeability in high-grade and low-grade brain tumors using dynamic susceptibility contrast MR imaging*. American Journal of Roentgenology, vol. 178, no. 3, pages 711–716, 2002. (Cited on page 35.)
- [Purcell 1946] E.M. Purcell, H.C.O. Torrey and R.V. Pound. *Resonance absorption by nuclear magnetic moments in a solid*. Physical review, vol. 69, no. 1-2, page 37, 1946. (Cited on page 23.)
- [Rector 1991] Nowlan W.A. Kay S. Rector A.L. *et al.* *Foundations for an electronic medical record*. Methods Inf Med, vol. 30, no. 3, pages 179–186, 1991. (Cited on page 124.)
- [Rodriguez 2012] J.D. Rodriguez, A. Perez, D. Arteta, D. Tejedor and J.A. Lozano. *Using multidimensional bayesian network classifiers to assist the treatment of multiple sclerosis*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pages 1705–1715, 2012. (Cited on page 156.)
- [Ron 2003] M.A. Ron and T.W. Robbins. Disorders of brain and mind, volume 2. Cambridge University Press, 2003. (Cited on page 26.)
- [Rosenkrantz 2015] A.B. Rosenkrantz, M. Mendiratta-Lala, B. JJ. Bartholmai, D. Ganeshan, R.G. Abramson, K.R. Burton, J.Y.U. John-Paul, E.M. Scalzetti, T.E. Yankeelov, R.M. Subramaniam *et al.* *Clinical utility of quantitative imaging*. Academic radiology, vol. 22, no. 1, pages 33–49, 2015. (Cited on page 32.)
- [Rosse 2003] C. Rosse and J.L.V. Mejino Jr. *A reference ontology for biomedical informatics : the Foundational Model of Anatomy*. Journal of biomedical informatics, vol. 36, no. 6, pages 478–500, 2003. (Cited on pages 76, 84, 99 and 104.)
- [RSNA 2007] RSNA. *Quantitative Imaging Biomarkers Alliance*. <https://www.rsna.org/QIBA.aspx>, [online] accessed 2015-06-18, 2007. (Cited on pages 6, 39 and 42.)
- [Rubin 2008a] D.L. Rubin. *Creating and curating a terminology for radiology : ontology modeling and analysis*. Journal of digital imaging, vol. 21, no. 4, pages 355–362, 2008. (Cited on page 62.)
- [Rubin 2008b] D.L. Rubin, C. Rodriguez, P. Shah and C.F. Beaulieu. *iPad : Semantic annotation and markup of radiological images*. In AMIA, 2008. (Cited on pages 50 and 68.)
- [Rubin 2014] D.L. Rubin, D. Willrett, M.J. O'Connor, C. Hage, C. Kurtz and D.A. Moreira. *Automated tracking of quantitative assessments of tumor burden in clinical trials*. Translational oncology, vol. 7, no. 1, pages 23–35, 2014. (Cited on pages 68, 124 and 143.)

- [Rudin 2003] M. Rudin and R. Weissleder. *Molecular imaging in drug discovery and development*. Nature reviews Drug discovery, vol. 2, no. 2, pages 123–131, 2003. (Cited on page 30.)
- [Ruttenberg 2007] A. Ruttenberg, T. Clark, W. Bug, M. Samwald, O. Bodenreider, H. Chen, D. Doherty, K. Forsberg, Y. Gao, V. Kashyap *et al.* *Advancing translational research with the Semantic Web*. BMC bioinformatics, vol. 8, no. Suppl 3, page S2, 2007. (Cited on page 70.)
- [Sanderson 2013] R. Sanderson, P. Ciccarese, H. Van de Sompel, S. Bradshaw, D. Brickley, L.J.G. Castro, T. Clark, T. Cole, P. Desenne, A. Gerber *et al.* *Open annotation data model*. W3C community draft, 2013. (Cited on page 76.)
- [Schaefer 2000] P.W. Schaefer, P.E. Grant and R.G. Gonzalez. *Diffusion-weighted MR imaging of the brain 1*. Radiology, vol. 217, no. 2, pages 331–345, 2000. (Cited on page 26.)
- [Scheuermann 2009] R.H. Scheuermann, W. Ceusters and B. Smith. *Toward an ontological treatment of disease and diagnosis*. Summit on translational bioinformatics, vol. 2009, page 116, 2009. (Cited on pages 8, 107 and 108.)
- [Schmainda 2012] K.M. Schmainda. *Diffusion-weighted MRI as a biomarker for treatment response in glioma*. CNS oncology, vol. 1, no. 2, pages 169–180, 2012. (Cited on page 26.)
- [Schriml 2011] L.M. Schriml, C. Arze, S. Nadendla, Y.W. Chang, M. Mazaitis, V. Felix, G. Feng and W.A. Kibbe. *Disease Ontology : a backbone for disease semantic integration*. Nucleic acids research, vol. 40, no. D1, pages D940–D946, 2011. (Cited on page 100.)
- [Schulz 2009] S. Schulz, H. Stenzhorn, M. Boeker and B. Smith. *Strengths and limitations of formal ontologies in the biomedical domain*. Revista electronica de comunicacao, informacao & inovacao em saude : RECIIS, vol. 3, no. 1, page 31, 2009. (Cited on page 70.)
- [Seifert 2010] S. Seifert, M. Kelm, M. Moeller, S. Mukherjee, A. Cavallaro, M. Huber and D. Comaniciu. *Semantic annotation of medical images*. In SPIE medical imaging, pages 762808–762808. International Society for Optics and Photonics, 2010. (Cited on pages 7, 84 and 86.)
- [Sirin 2007] E. Sirin, B. Parsia, B.C. Grau, A. Kalyanpur and Y. Katz. *Pellet : A practical owl-dl reasoner*. Web Semantics : science, services and agents on the World Wide Web, vol. 5, no. 2, pages 51–53, 2007. (Cited on page 129.)
- [Smith 2005a] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A.L. Rector and C. Rosse. *Relations in biomedical ontologies*. Genome biology, vol. 6, no. 5, page R46, 2005. (Cited on pages 75 and 97.)
- [Smith 2005b] B. Smith, A. Kumar and T. Bittner. *Basic formal ontology for bioinformatics*. Journal of Information Systems, pages 1–16, 2005. (Cited on pages 75 and 94.)

- [Smith 2006a] B. Smith. *From concepts to clinical reality : an essay on the benchmarking of biomedical terminologies*. Journal of biomedical informatics, vol. 39, no. 3, pages 288–298, 2006. (Cited on pages 98, 124 and 125.)
- [Smith 2006b] B. Smith, W. Kusnierczyk, D. Schober and W. Ceusters. *Towards a reference terminology for ontology research and development in the biomedical domain*. In KR-MED, volume 2006, pages 57–66, 2006. (Cited on pages 125, 142 and 143.)
- [Smith 2007] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungallet *al.* *The OBO Foundry : coordinated evolution of ontologies to support biomedical data integration*. Nature biotechnology, vol. 25, no. 11, pages 1251–1255, 2007. (Cited on pages 75 and 98.)
- [Smith 2012] B. Smith. *Classifying processes : an essay in applied ontology*. Ratio, vol. 25, no. 4, pages 463–488, 2012. (Cited on pages 7 and 97.)
- [Smith 2014] E.T.S. Smith. *Clinical applications of imaging biomarkers. Part 1. The neuroradiologist’s perspective*. The British journal of radiology, 2014. (Cited on page 32.)
- [Smith 2015] B. Smith, S. Arabandi, M. Brochhausen, M. Calhoun, P. Ciccarese, S. Doyle, B. Gibaud, I. Goldberg, C.E. Kahn Jr, J. Overtonet *al.* *Biomedical imaging ontologies : A survey and proposal for future work*. Journal of pathology informatics, vol. 6, 2015. (Cited on pages 7 and 101.)
- [SNOMED 2015] SNOMED. *SNOMED CT*. <http://www.snomed.org/snomedct/documents.html>, [online] accessed 2014-11-24, 2015. (Cited on page 61.)
- [Starkbaum 2014] J. Starkbaum, H. Gottweis, U. Gottweis, C. Kleiser, J. Linseisen, C. Meisinger, P. Kamtsiuris, S. Moebus, K.H. Joeckel, S. Boermet *al.* *Public perceptions of cohort studies and biobanks in Germany*. Biopreservation and biobanking, vol. 12, no. 2, pages 121–130, 2014. (Cited on page 48.)
- [Sullivan 2015] D.C. Sullivan, N.A. Obuchowski, L.G. Kessler, D.L. Raunig, C. Gattsonis, E.P. Huang, M. Kondratovich, L.M. McShane, A.P. Reeves, D.P. Barboriak *et al.* *Metrology standards for quantitative imaging biomarkers*. Radiology, vol. 277, no. 3, pages 813–825, 2015. (Cited on page 39.)
- [TCGA-Research-Group 2014] TCGA-Research-Group. *The Cancer Genome Atlas Glioblastoma Multiforme data collection*. <https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM>, 2014. (Cited on pages 11, 117 and 121.)
- [Temal 2008] L. Temal, M. Dojat, G. Kassel and B. Gibaud. *Towards an ontology for sharing medical images and regions of interest in neuroimaging*. Journal of Biomedical Informatics, vol. 41, no. 5, pages 766–778, 2008. (Cited on pages 7 and 79.)
- [Tofts 1999] P.S. Tofts, G. Brix, D.L. Buckley, J.L. Evelhoch, E. Henderson, M.V. Knopp, H.B.W. Larsson, T.Y. Lee, N.A. Mayr, G.J.M. Parker *et al.* *Estimating kinetic parameters from dynamic contrast-enhanced T 1-weighted MRI*

- of a diffusable tracer : standardized quantities and symbols*. Journal of Magnetic Resonance Imaging, vol. 10, no. 3, pages 223–232, 1999. (Cited on pages 27 and 35.)
- [Tofts 2010] P.S. Tofts. *T1-weighted DCE imaging concepts : modelling, acquisition and analysis*. signal, vol. 500, no. 450, page 400, 2010. (Cited on page 27.)
- [Velazquez 2015] E.R. Velazquez, R. Meier, W.D. Dunn Jr, B. Alexander, R. Wiest, S. Bauer, D.A. Gutman, M. Reyes and H.J.W.L Aerts. *Fully automatic GBM segmentation in the TCGA-GBM dataset : Prognosis and correlation with VASARI features*. Scientific reports, vol. 5, 2015. (Cited on page 143.)
- [Verma 2011] S. Verma, A. Rajesh, H. Morales, L. Lemen, G. Bills, M. Delworth, K. Gaitonde, J. Ying, R. Samartunga and M. Lamba. *Assessment of aggressiveness of prostate cancer : correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy*. American Journal of Roentgenology, vol. 196, no. 2, pages 374–381, 2011. (Cited on page 26.)
- [VIM 2012] VIM. *International vocabulary of metrology - basic and general concepts and associated terms*, 2012. (Cited on page 92.)
- [Vos 2013] E.K. Vos, G.J.S. Litjens, T. Kobus, T. Hambrock, C.A. Hulsbergen-van de Kaa, J.O. Barentsz, H.J. Huisman and T.W.J. Scheenen. *Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced magnetic resonance imaging at 3 T*. European urology, vol. 64, no. 3, pages 448–455, 2013. (Cited on page 27.)
- [Wahl 1993] R.L. Wahl, K. Zasadny, M. Helvie, G.D. Hutchins, B. Weber and R. Cody. *Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography : initial evaluation*. Journal of Clinical Oncology, vol. 11, no. 11, pages 2101–2111, 1993. (Cited on pages 30 and 32.)
- [Wahl 2009] R.L. Wahl, H. Jacene, Y. Kasamon and M.A. Lodge. *From RECIST to PERCIST : evolving considerations for PET response criteria in solid tumors*. Journal of nuclear medicine, vol. 50, no. Suppl 1, pages 122S–150S, 2009. (Cited on pages 32 and 35.)
- [Waterton 2008] J.C. Waterton. *What imaging biomarkers are and how they are used*. 2008. (Cited on page 45.)
- [Weber 2006] W.A. Weber. *Positron emission tomography as an imaging biomarker*. Journal of Clinical Oncology, vol. 24, no. 20, pages 3282–3292, 2006. (Cited on page 30.)
- [Weber 2009] W.A. Weber. *Assessing tumor response to therapy*. Journal of nuclear medicine, vol. 50, no. Suppl 1, pages 1S–10S, 2009. (Cited on page 34.)
- [Weng 2010] C. Weng, S.W. Tu, I. Sim and R. Richesson. *Formal representation of eligibility criteria : a literature review*. Journal of biomedical informatics, vol. 43, no. 3, pages 451–467, 2010. (Cited on page 156.)
- [West 2014] C.M.L. West and N. Charnley. *The potential of PET to increase understanding of the biological basis of tumour and normal tissue response to radiotherapy*. The British Journal of Radiology, 2014. (Cited on page 30.)

- [Wintermark 2008] M. Wintermark, G.W. Albers, A.V. Alexandrov, J.R. Alger, R. Bammer, J.C. Baron, S. Davis, B.M. Demaerschalk, C.P. Derdeyn, G.A. Donnane *et al.* *Acute stroke imaging research roadmap*. American Journal of Neuroradiology, vol. 29, no. 5, pages e23–e30, 2008. (Cited on page 43.)
- [Woodbridge 2013] M. Woodbridge, G. Fagiolo and D.P. O'Regan. *MRIdb : medical image management for biobank research*. Journal of digital imaging, vol. 26, no. 5, pages 886–890, 2013. (Cited on page 48.)
- [Woodfield 2010] C.A. Woodfield, G.A. Tung, D.J. Grand, J.A. Pezzullo, J.T. Machan and J.F. Renzulli. *Diffusion-weighted MRI of peripheral zone prostate cancer : comparison of tumor apparent diffusion coefficient with Gleason score and percentage of tumor on core biopsy*. American Journal of Roentgenology, vol. 194, no. 4, pages W316–W322, 2010. (Cited on page 26.)
- [Xiang 2010] Z. Xiang, M. Courtot, R.R. Brinkman, A. Ruttenberg and Y. He. *OntoFox : web-based support for ontology reuse*. BMC research notes, vol. 3, no. 1, page 175, 2010. (Cited on page 111.)
- [Yankeelov 2007] T.E. Yankeelov, M. Lepage, A. Chakravarthy, E.E. Broome, K.J. Niermann, M.C. Kelley, I. Meszoely, I.A. Mayer, C.R. Herman, K. McManuset *et al.* *Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer : initial results*. Magnetic resonance imaging, vol. 25, no. 1, pages 1–13, 2007. (Cited on pages 26 and 35.)
- [Zahra 2007] M.A. Zahra, K.G. Hollingsworth, E. Sala, D.J. Lomas and L.T. Tan. *Dynamic contrast-enhanced MRI as a predictor of tumour response to radiotherapy*. The lancet oncology, vol. 8, no. 1, pages 63–74, 2007. (Cited on page 27.)
- [Zarow 2005] C. Zarow, H.V. Vinters, W.G. Ellis, M.W. Weiner, D. Mungas, L. White and H.C. Chui. *Correlates of hippocampal neuron number in Alzheimer's disease and ischemic vascular dementia*. Annals of neurology, vol. 57, no. 6, pages 896–903, 2005. (Cited on page 25.)
- [Zhao 2010] B. Zhao, G.R. Oxnard, C.S. Moskowitz, M.G. Kris, W. Pao, P. Guo, V.M. Rusch, M. Ladanyi, N.A. Rizvi and L.H. Schwartz. *A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development*. Clinical Cancer Research, vol. 16, no. 18, pages 4647–4653, 2010. (Cited on page 23.)
- [Zheng 2008] H.T. Zheng, B.Y. Kang and H.G. Kim. *An ontology-based bayesian network approach for representing uncertainty in clinical practice guidelines*. In Uncertainty Reasoning for the Semantic Web I, pages 161–173. Springer, 2008. (Cited on page 156.)
- [Zussman 2011] B. Zussman, P. Jabbour, K. Talekar, R. Gorniak and A.E. Flanders. *Sources of variability in computed tomography perfusion : implications for acute stroke management*. Neurosurgical focus, vol. 30, no. 6, page E8, 2011. (Cited on page 43.)