UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Un wiki sémantique pour la gestion des connaissances décisionnelles – Application à la cancérologie

THÈSE

présentée et soutenue publiquement le 28 juin 2013

pour l'obtention du

Doctorat de l'Université de Lorraine

(Mention informatique)

par

Thomas MEILENDER

Composition du jury

Rapporteurs :	Isabelle Bichindaritz	Assistant professor, State University of New York, Oswego
	Marie-Christine Jaulent	Directeur de recherche, INSERM, Paris
Examinateurs :	Sylvie Després	Professeur, Université Paris 13
	Monique Grandbastien	Professeur, Université de Lorraine
Directeur :	Jean Lieber	Maître de conférences (H.D.R), Université de Lorraine
Co-directeur :	Nicolas Jay	Maître de conférences, Université de Lorraine

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Loria
Laboratoire lorrain de recherche en informatique et ses applications

Mis en page avec la classe thloria.

Remerciements

J'exprime tout d'abord mes remerciements à Isabelle Bichindaritz et Marie-Christine Jaulent qui ont accepté d'être les rapporteurs de cette thèse et qui par leurs remarques m'ont permis d'améliorer ce mémoire. Merci également à Sylvie Després ainsi qu'à Monique Grandbastien qui ont accepté d'être respectivement examinateur et présidente du jury.

Je tiens également à exprimer mes remerciements et ma gratitude à Jean Lieber, mon directeur de thèse, pour sa compétence, sa pédagogie et son soutien. Ces années ont été riches en enseignements grâce à toi, tant au niveau personnel que professionnel. Ce fut un honneur pour moi de travailler avec ce rat-taupe glabre de la recherche.

Merci également à Nicolas Jay de m'avoir encadré et notamment de m'avoir apporté ses lumières sur l'informatique médical.

De nombreuses personnes ont participé au projet KASIMIR ces dernières années. Je souhaite particulièrement remercier Fabien Palomarès d'A2ZI d'avoir cru au projet et d'avoir financer ce travail et Gilles Herengt d'Oncolor pour son aide, ses idées et son soutien. Merci également aux personnels d'A2ZI et d'Oncolor pour leur collaboration.

D'autres collaborateurs ont mis leur talent au service du projet. Un grand merci à eux, notamment à Aline Zimmer, Fadi Badra, Mathieu d'Aquin, Anh-Djuy Tran, Valentien Jeannot, Jonathan Leeuwen et Anthony Kirkpatrick.

Merci à Amedeo Napoli de m'avoir accueilli à Orpailleur, ainsi qu'à tous les autres membres de l'équipe que j'ai eu la chance de côtoyer. Merci à tous les amis du Loria pour les discussions et les bons moments passés ensemble : Stéphane, Yves, Hatem, Nizar, Mehdi, Julien, Saïna, Laura, Alice, Nicolas, Florent, Yesmine, Hana, Elias, Valmi, Emmanuelle... J'en oublie probablement parmi les meilleurs.

Je n'aurai probablement pas mené aussi loin mes études sans le soutien de ma famille. Merci à ma Maman d'avoir toujours été là, à Nick et Marie, à PoJo, à Aurore et Olivier (tous mes vœux de bonheur !) et à ma sœur de de cœur Alex. Merci à mes filleules, Katy et Orane, qui sans le savoir ont été d'un soutien moral extraordinaire. Longue et belle vie à Camille et Louise, et à celui qui va venir. Et merci également d'avoir été là à Papy et Manou, Xavier, Fabienne et Yann, les cousins et cousines. Sans oublier un grand merci à la famille Bendaoud de m'avoir si chaleureusement accueilli.

Pour ma merveilleuse Hanna, que ta vie soit toujours bonheur et sourire.

Et surtout, merci Rokia pour ton soutien et pour cette superbe aventure que nous vivons tous les jours.

Pour mon épouse et pour ma fille,
les deux piliers de ma vie.

Table des matières

Table des figures

Liste des tableaux

Introduction générale

Ces vingt dernières années, l'explosion du nombre de sources documentaires sur l'Internet a conduit à ce paradoxe : si l'accès aux documents n'a jamais été aussi simple et rapide, leur quantité rend leur recherche plus complexe et plus coûteuse. Ainsi, les informations sont souvent mal exploitées à cause de leur hétérogénéité et de la difficulté de les localiser. En effet, la nature des documents publiés présente une grande variabilité. Pour optimiser l'exploitation de ces informations, de nouveaux outils sont nécessaires pour structurer et indexer les différentes ressources.

Parmi les différents documents disponibles se trouvent des guides de bonnes pratiques. Utiles dans différents domaines, ces documents contiennent des connaissances d'un type particulier, appelées connaissances décisionnelles. Destinées à décrire un processus de prise de décision, elles permettent d'associer à une situation particulière une recommandation dépendant du contexte décrit. Localiser et identifier de manière formelle ces connaissances permettraient de créer des bases de connaissances destinées à servir de support à des systèmes automatiques pour l'aide à la décision. Toutefois, la plupart du temps, ces connaissances sont éditées par des experts sur des supports statiques destinés à des êtres humains. Les informations restent alors incompréhensibles par des machines. Ce mode d'édition est dû au manque d'outils disponibles permettant de coupler l'édition et la formalisation des connaissances décisionnelles. La plupart des outils permettant l'édition formelle de connaissances sont assez complexes à appréhender car destinés à des ingénieurs des connaissances, spécialistes de ce type d'édition. De plus, peu fournissent un résultat compréhensible et exploitable par un utilisateur classique. Pour qu'ils puissent être adoptés par des experts aux compétences restreintes en ingénierie des connaissances, ces outils doivent donc être adaptés et simplifiés afin de limiter un accroissement de la masse de travail nécessaire à la formalisation des connaissances.

Contexte applicatif. Dans de nombreux domaines, et en particulier en médecine, l'accès aux documents et leur utilisation sont des enjeux majeurs pour le grand public et les professionnels. Ainsi, l'information médicale devient de plus en plus disponible et accessible. L'Internet propose de nombreux sites pour informer les praticiens et pour améliorer la communication. Les sites de référencement des publications médicales telles que PubMed [NCBI, 2013b] et CISMeF [CISMeF, 2013] vont dans ce sens. À titre d'exemple, PubMed indexe plus de 21 millions de références bibliographiques issues de la littérature médicale. La profusion de ressources et les nouveaux besoins engendrés par l'explosion des nouvelles technologies de l'information et de la communication ont conduit à la création d'outils pour la structuration des informations, tels que les thésaurus médicaux qui reposent sur le vocabulaire spécialisé du domaine. On peut citer par exemple SNOMED pour le codage des informations cliniques, MeSH pour l'indexation des ressources documentaires médicales ou la CCAM pour le codage médico-économique.

Parmi les documents médicaux améliorant la prise en charge du patient, les recommandations

de pratique clinique (RPC[1]) sont définies par la HAS[2] comme des propositions développées méthodiquement pour aider le praticien et le patient à rechercher les soins les plus appropriés pour des caractéristiques cliniques données [Haute Autorité de la Santé, 2010]. Leur but est d'améliorer la qualité des soins et de réduire les coûts en s'appuyant sur des recommandations élaborées par des organismes certifiés tels que la HAS ou les sociétés savantes. Les RPC sont le plus souvent disponibles sous forme de documents papiers ou dans un format électronique imprimable du type PDF, ce qui les rend difficiles à utiliser dans un contexte clinique. Idéalement, les RPC reposent sur les principes de la médecine fondée sur les faits. Toutefois, devant la difficulté à réunir les meilleures données cliniques, la plupart des RPC sont construites à partir de consensus d'experts médicaux, ce qui induit la nécessité d'un travail collaboratif de leur part. Ces consensus ne sont pas toujours faciles à obtenir mais doivent pourtant être trouvés à chaque ajout de connaissances. De plus, afin de rendre les RPC compréhensibles au public le plus large possible, un travail de mise en forme fastidieux est nécessaire. En effet, l'édition de ces informations peut prendre plusieurs formes : des paragraphes de textes, des schémas explicatifs, des nomenclatures ou des arbres de décision.

Le but des systèmes d'aide à la décision (SAD) est de fournir la recommandation la plus adéquate dans le cadre d'une prise de décision. Ainsi, plusieurs SAD prennent en compte les informations contenues dans les RPC et cherchent à mettre en œuvre les connaissances recueilllies, répondant à un besoin de plus en plus exprimé par les professionnels de santé. Pour être efficaces, les SAD doivent être alimentés en connaissances formelles issues généralement du savoir des experts du domaine. Or, les connaissances sous forme documentaire ne peuvent pas être exploitées par les SAD : ceux-ci nécessitent des connaissances formelles, manipulables par des moteurs d'inférences. Toutefois, la capitalisation des connaissances est une étape très coûteuse en temps d'experts du domaine et d'ingénieurs des connaissances.

Cette capitalisation des connaissances décisionnelles fait référence à l'ingénierie des connaissances, qui se définit comme une discipline chargée de l'intégration des connaissances dans les systèmes informatisés dans le but de résoudre des problèmes complexes nécessitant un haut degré d'expertise humaine [Feigenbaum et McCorduck, 1983]. Elle est donc impliquée dans la construction, la maintenance et le développement des systèmes à base de connaissances. Actuellement, le cadre technique de la formalisation des connaissances est fourni par le Web sémantique. Dans le Web classique, la recherche et l'agrégation des informations se font par des humains qui interprètent les résultats issus des moteurs de recherche. Puisque les pages sont réalisées à l'intention d'humains, il est actuellement impossible pour un processus automatique d'en saisir le sens. Le Web sémantique est un projet dont le but est d'améliorer la présentation des documents en ligne afin de les rendre compréhensibles par des processus automatiques, permettant ainsi à des machines de chercher, combiner, gérer et produire des connaissances. Sous l'égide du W3C, des standards sont apparus pour exprimer de manière formelle les connaissances, tels RDF [Swick et Lassila, 1999]) et OWL [Dean *et al.*, 2002], et pour les interroger, tel que le langage de requêtes SPARQL [Prud'hommeaux et Seaborne, 2008a].

Problématique. D'une part, les SAD reposent sur des bases de connaissances décisionnelles nécessitant un développement coûteux pour être fonctionnelles et, d'autre part, des experts travaillent à la rédaction de RPC contenant ces mêmes connaissances. Il apparaît qu'idéalement, s'il existait des outils permettant de structurer les connaissances lors du processus d'édition, le gain serait important pour les SAD. Le but de cette thèse est donc de trouver des méthodes

1. La liste des abréviations utilisées est présentée page 138.
2. Haute Autorité de Santé.

qui permettraient de factoriser les moyens mis en œuvre pour d'une part l'édition des RPC et d'autre part la création et la maintenance de bases de connaissances.

Actuellement, l'avancée technologique des wikis sémantiques rend possible une telle édition. Les wikis sont des sites Web permettant la création et l'édition collaborative de contenus de manière simple [Leuf et Cunningham, 2001]. Ils reposent généralement sur un ensemble de pages éditables, organisées en catégories et reliées par des liens hypertextes et dont l'historique peut être contrôlé grâce à des mécanismes de suivi des modifications. Un wiki sémantique est un wiki dont les capacités sont augmentées par l'utilisation des technologies du Web sémantique. Ils permettent de formaliser le sens des articles par l'ajout de metadonnées sur les pages et la caractérisation des liens entre ces pages [Krötzsch *et al.*, 2007b]. Les informations formalisées deviennent donc exploitables par une machine, à travers des processus de raisonnement artificiel. Toutefois, les wikis sémantiques ne permettent pas la gestion des connaissances décisionnelles. Il convient donc de déterminer des méthodes et des techniques pour leur prise en compte qui soient assez simples pour être utilisées par des experts médicaux.

Organisation du Mémoire. La première partie présente le projet de recherche KASIMIR qui vise à fournir des outils pour la gestion des connaissances décisionnelles en oncologie. Il réunit des acteurs issus de différents domaines tels que l'ingénierie informatique (la SARL A2ZI), la médecine (le réseau régional de cancérologie en Lorraine Oncolor) et la recherche universitaire (le laboratoire Loria). Sur son site Internet statique, Oncolor publie 146 RPC en oncologie, écrites en HTML. Afin d'en faciliter la création, la maintenance et la publication, Oncolor a exprimé le besoin d'outils collaboratifs. Par ailleurs, il serait intéressant pour les systèmes d'aide à la décision que les connaissances contenues dans les RPC soient formalisées et mises à la disposition des systèmes sémantiques, en particulier pour KASIMIR. Il conviendra alors de présenter les connaissances techniques mises en jeu, telles que le Web sémantique et les ressources médicales à disposition.

La deuxième partie présentera les wikis sémantiques. Après les avoir définis, nous nous intéresserons aux approches ayant guidé la conception de ces gestionnaires de contenu particuliers ainsi qu'aux enjeux liés à ce type de système, en particulier la manière dont les annotations sont exploitées. Les principaux moteurs de wiki sémantique seront ensuite présentés, ainsi que des exemples d'application.

La création d'un wiki sémantique pour les RPC en oncologie est abordée dans la troisième partie. Reposant sur les versions HTML éditées par Oncolor, OncologiK [3] a nécessité la mise en place d'un processus de migration d'un site statique de type Web 1.0 vers une solution du Web social sémantique. Les bénéfices attendus de cette migration sont de deux ordres : premièrement, le travail d'édition est simplifié par l'utilisation des wikis et, deuxièmement, les technologies sémantiques permettent la création de services supplémentaires utilisant des ressources externes telles que les terminologies médicales et les sites Web des publications médicales. La migration implique la nécessité de prendre en compte le contenu structuré et non structuré, mais aussi, lorsque c'est possible, d'inclure des outils dédiés afin de faciliter la formalisation des données. Parfois, le développement de nouvelles extensions est nécessaire.

Les documents importés comprennent de nombreuses connaissances décisionnelles représentées sous forme d'arbres de décision. La plupart du temps, les arbres de décision peuvent être considérés comme des structures à partir desquelles un sens peut être extrait. De nombreux formalismes ont déjà été proposés en informatique médicale afin de les représenter et de proposer des guides de bonnes pratiques informatisés. Toutefois, aucun de ces formalismes ne s'est

3. `http://www.oncologik.fr`

imposé comme un standard universel. L'objet de la quatrième partie est de présenter les connaissances décisionnelles dans le contexte des référentiels médicaux, ainsi que de passer en revue les principaux formalismes et outils qui permettent de les exprimer.

Le traitement des arbres de décision est le sujet de la cinquième partie. Il apparaît que leur création et leur mise à jour pourraient être simplifiées s'il existait un éditeur d'arbres de décision en ligne, compatible avec le wiki sémantique. KcatoS, un éditeur répondant à ces besoins, est proposé. Il dispose d'un algorithme d'export qui permet de transformer les arbres de décision en OWL. Ainsi, les ontologies créées sont mises à disposition d'autres applications du Web sémantique.

De nombreuses perspectives sont évoquées au cours de ce travail, aussi bien au niveau des wikis sémantiques que de la représentation des connaissances décisionnelles. Plusieurs seront développées dans la sixième partie et, enfin, une conclusion cloture cette thèse.

Chapitre 1

Contexte applicatif et scientifique

Sommaire

Ce travail s'inscrit dans le cadre du projet KASIMIR, qui s'intéresse à la gestion des connaissances décisionnelles en oncologie. La section 1.1 en expose les objectifs, les acteurs et les besoins. Le projet s'appuie notamment sur les guides de bonnes pratiques cliniques, dont les buts et les méthodes de rédaction sont décrits dans la section 1.2. Le contexte technique du projet est fourni par le Web sémantique, présenté dans la section 1.3. Développé depuis une quinzaine d'années, le Web sémantique propose des applications dans de nombreux domaines, en particulier dans celui de l'informatique médicale. Plusieurs de ces applications sont présentées dans la section 1.4.

1.1 Le projet KASIMIR

1.1.1 KASIMIR : objectifs et acteurs

Démarré en 1997, le projet de recherches KASIMIR [KASIMIR, 2013] est un projet informel qui vise à proposer des méthodes et à créer des outils logiciels pour l'aide à la décision et, plus généralement, pour la gestion des connaissances décisionnelles en oncologie [Lieber, 2006]. Les connaissances étudiées sont issues des guides de bonnes pratiques cliniques [4] (GBPC) en cancérologie, documents composés de l'ensemble des RPC nécessaires à la prise en charge des patients pour une pathologie. L'objectif est de formaliser ces connaissances de manière à les rendre utilisables par des systèmes intelligents en fournissant les outils pour leur maintenance et leur mise à jour.

De nombreuses personnes, issues de diverses disciplines, ont contribué au projet KASIMIR, parmi ceux-ci des experts en cancérologie du Centre Alexis Vautrin (désormais Institut de Cancérologie de Lorraine, Vandœuvre-lès-Nancy), des informaticiens de l'association Hermès (désormais GCS Télésanté Lorraine, hébergement des réseaux mutualisés en santé) et des chercheurs du laboratoire d'ergonomie du CNAM (Paris). Le projet a réuni principalement trois contributeurs ces dernières années :

- l'équipe de recherche Orpailleur [Orpailleur, 2013] du Loria, qui étudie l'extraction et la représentation de connaissances,
- l'entreprise A2ZI [A2ZI, 2013], implantée à Commercy, spécialisée dans l'ingénierie du Web en particulier dans le domaine médical,
- le réseau de santé Oncolor [Oncolor, 2013], dont l'objectif est d'améliorer la prise en charge des patients atteints de cancer en région Lorraine.

Le projet a pour origine l'observation de l'utilisation des GBPC par les praticiens. Ces guides contiennent les informations pour la prise en charge des patients dans des cas « standard ». D'un point de vue informatique, ces informations peuvent être formalisées sous forme de connaissances décisionnelles. Pour tous les cas présentant des variations, une adaptation spécifique des recommandations doit être faite. Par exemple, dans le cadre du cancer du sein, les GBPC ne sont strictement applicables que dans 60 à 70% des cas [Lieber *et al.*, 2008]. Dans les autres cas, le protocole de soins standard doit alors être adapté par le praticien, ce qui signifie que le GBPC ne sera pas appliqué tel quel mais modifié en fonction du cas clinique. Ces modifications sont décidées collégialement lors de réunions de concertation pluri-disciplinaires (RCP) regroupant des experts issus des différentes disciplines de la cancérologie. Si elles sont applicables à d'autres

4. Dans le cadre du projet KASIMIR, les GBPC sont communément appelés des référentiels, le terme emprunté à la terminologie d'Oncolor.

cas, ces modifications peuvent être à l'origine d'une mise à jour du GBPC. L'objectif du projet KASIMIR est alors de proposer des modèles informatiques pouvant rendre compte de ces adaptations et de ces évolutions.

1.1.2 Vers un portail sémantique

Dans le cadre du projet, plusieurs pistes ont été explorées pour l'acquisition, la représentation et l'exploitation des connaissances décisionnelles en oncologie. Ainsi, des principes d'acquisition des connaissances pour le raisonnement à partir de cas ont été proposés [d'Aquin *et al.*, 2006a], ainsi que des études de modélisation avec des logiques floues [d'Aquin *et al.*, 2006b] ou l'exploitation des règles d'adaptation pour la création de nouvelles règles [d'Aquin *et al.*, 2007]. Ces études ont mis en avant le besoin d'outils logiciels pour capitaliser les connaissances et de standards pour les représenter. En effet, les premières versions des systèmes KASIMIR reposaient sur des formats *ad hoc*, ne permettant qu'une faible interopérabilité avec des outils externes.

De ce bilan est né le besoin d'un portail sémantique en oncologie [d'Aquin *et al.*, 2005]. En s'appuyant sur les standards du Web sémantique, le but du portail est de fournir un environnement intelligent pour la gestion des connaissances décisionnelles et l'aide à la décision. Plusieurs outils ont déjà été proposés dans le cadre de ce portail, certains sont présentés dans la section 1.4.3.

1.2 Les guides de bonnes pratiques cliniques

Le projet KASIMIR repose donc sur les connaissances décisionnelles issues des guides de bonnes pratiques cliniques. Cette section présente ces guides, en particulier ceux édités par le réseau Oncolor qui servent de support à ce travail.

1.2.1 Définition

Au cours des deux dernières décennies, les GBPC sont devenus des éléments familiers dans la pratique clinique. Hoyt [1997] les définit comme « des instructions développées méthodologiquement afin d'aider les praticiens et les patients sur les décisions de soins de santé appropriés pour des circonstances cliniques spécifiques. » [5]. Généralement élaborés par des sociétés savantes ou des agences nationales, ils proposent des instructions concises et valides sur les pratiques cliniques telles que les diagnostics, les examens à prescrire ou les temps d'observation. À travers leur utilisation, le but recherché est de standardiser les soins cliniques en fonction de l'état de l'art et des données acquises de la science afin d'améliorer la qualité des soins et la prise en charge des patients tout en réduisant le coût des soins [Georg, 2006]. Ces documents peuvent être distribués en version papier ou en version électronique souvent disponible sur l'Internet.

L'utilisation des GBPC par les professionnels en médecine présente de nombreux avantages détaillés par Woolf *et al.* [1999]. En faisant la promotion des bonnes pratiques et en décourageant les pratiques inefficaces, ils permettent d'améliorer la qualité des soins. De plus, les guides compréhensibles par le grand public informent les patients sur les actes qui doivent être effectués. Mieux informés, les patients peuvent faire des choix plus éclairés en fonction de leurs besoins personnels et de leurs préférences. Plus globalement, les GBPC peuvent influencer les politiques

5. « *Clinical practice guidelines are systematically developed statements to assist practitioners' and patients' decisions about appropriate health care for specific clinical circumstances.* ».

de santé publique en mettant en avant des problèmes de santé sous-estimés, des services cliniques méconnus ou en identifiant des populations à risque. Du point de vue des praticiens, les guides sont plutôt vus comme des outils de formation permanente. Ainsi, ils fournissent des recommandations explicites aux cliniciens qui ne sont pas sûrs quant à la manière de procéder et modifient les croyances des praticiens habitués à des pratiques dépassées. Enfin, ils peuvent également être une référence pour les évaluations médicales, en les comparant avec les pratiques des professionnels ou des hôpitaux.

Les contenus des GBPC sont définis selon trois méthodes [Hoyt, 1997] :

- les consensus d'experts : cette méthode repose sur des recommandations ayant obtenu l'unanimité des experts médicaux. Largement critiquée et de moins en moins utilisée, cette technique ne repose pas sur un lien explicite entre la recommandation et la preuve de la qualité du traitement.
- la médecine factuelle (*evidence-based medicine*) : elle propose d'argumenter les décisions par des preuves scientifiques telles que des essais cliniques ou des études. Toutefois, dans de nombreux cas, des recommandations ne peuvent être faites à cause de l'absence de preuve acceptable.
- les méthodes explicites : cette approche propose de prendre en compte les preuves scientifiques et l'opinion des experts. D'un point de vue pratique, les bénéfices, dangers et coûts de chaque intervention sont mis en évidence de manière à fournir le plus d'éléments possible aux praticiens et aux patients.

Cependant, certaines limites des GBPC ont déjà été mises en avant [Woolf *et al.*, 1999]. La principale concerne les recommandations : elles peuvent être fausses à cause d'erreurs humaines, de l'utilisation de connaissances médicales dépassées, d'une mauvaise interprétation ou du manque de preuve scientifique. De plus, une recommandation peut ne pas être applicable dans certains cas, sans qu'une adaptation ne soit possible. Pour certains GBPC, les recommandations peuvent être influencées par l'opinion et l'expérience des rédacteurs du guide. Il se peut également que les besoins du patient ne soient pas la priorité des rédacteurs ou de l'organisme finançant sa rédaction, et que les recommandations soient guidées par d'autres facteurs tels que le contrôle des coûts ou la protection d'intérêts spéciaux. Enfin, même lorsqu'ils sont corrects, les praticiens trouvent souvent des inconvénients à leur utilisation [Cabana *et al.*, 1999; Lugtenberg *et al.*, 2011]. Outre le manque d'applicabilité dans certains cas, de nombreux praticiens avouent avoir du mal à changer leurs habitudes et trouvent que l'utilisation des GBPC dans un contexte clinique prend trop de temps.

1.2.2 Création des guides de bonnes pratiques cliniques en oncologie

1.2.2.1 Description des GBPC Oncolor

Dans le cadre de ses activités, le réseau Oncolor coordonne la rédaction, l'édition et la publication de GBPC en oncologie. Ces documents synthétiques sont élaborés par des groupes de travail pluridisciplinaires composés de professionnels de santé membres du réseau. 146 guides ont déjà été édités par Oncolor, concernant les localisations de cancer, les types histologiques ou des situations de prise en charge telles que les soins de support ou les traitements. Destinés aussi bien aux usagers qu'à la formation des professionnels de la santé, les GBPC doivent être précis et didactiques. Pour cela, plusieurs formes de rédaction sont utilisées : des textes, des schémas, des tableaux ou des arbres de décision.

Les guides sont disponibles sous forme électronique sur le site Web d'Oncolor. Pour chaque guide, une version HTML et une version PDF imprimable sont proposées. Certains guides ont également été distribués au format papier, sous forme de petits livres mis en page sur le modèle des versions PDF. Le site Web d'Oncolor diffuse également des informations diverses sur les services de santé locaux et propose un accès à des outils professionnels. Il héberge également un thésaurus relatif à des produits pharmacologiques utilisés dans le traitement des cancers, auquel il est fréquemment fait référence dans les GBPC.

La plupart des GBPC sont construits selon la même structure. La première partie introduit le document avec quelques phrases qui expliquent les circonstances et les conditions d'utilisation du guide, ainsi que les traitements qui seront proposés. La partie suivante est une description textuelle des examens cliniques et paracliniques qui conduisent à la description de la situation standard de départ du GBPC. Ce point de départ est souvent une étape intermédiaire permettant de classer le patient selon les classifications internationales, représentées sous forme de tableaux simples. En fonction des résultats des classifications, des arbres de décision guident le lecteur à l'étape suivante où sont détaillées les recommandations médicales applicables. Ces recommandations sont disponibles en divers formats, tels que des publications médicales au format PDF ou des ressources distantes. Enfin, les GBPC se concluent par des conseils sur la surveillance médicale et parfois avec un lexique des termes spécifiques. La figure 1.1 montre des extraits de GBPC issus du site Web d'Oncolor.

1.2.2.2 Cycle de vie d'un GBPC : rédaction, maintenance et diffusion

Les GBPC édités par Oncolor sont construits selon les principes de la médecine factuelle ou, à défaut de preuve, grâce aux consensus d'experts. Comme dans tous les systèmes d'information médicaux, les informations contenues sont critiques. Elles font l'objet de révisions régulières. Ainsi, théoriquement, chaque GPBC doit être révisé au moins une fois tous les deux ans afin de le mettre à jour en fonction des données acquises de la science.

Pour chaque GBPC, un coordinateur est nommé au sein d'Oncolor pour assurer le suivi, la cohérence et la complétude du document. Deux types de contributeurs peuvent être identifiés :

- Les experts médicaux apportent leurs connaissances techniques. Ils sont réunis en comités, sous la supervision du coordinateur.
- Le personnel d'Oncolor gère la communication entre les membres des comités et assure la mise en page finale. Il vérifie également que les GBPC sont à jour et propose de nouveaux moyens pour favoriser leur diffusion.

La répartition des tâches et des rôles au sein d'Oncolor lors de la création et la mise à jour des GBPC a fait l'objet d'un étude détaillée dans le cadre du projet Kasimir [Zimmer, 2009].

Les documents sont écrits en respectant les principes de la certification HonCODE [HonCODE, 2013]. Le but de celle-ci est de garantir la qualité, l'objectivité et la transparence des informations en instaurant un code de déontologie conforme aux pratiques médicales. Entre autres choses, ce code impose d'identifier les responsables des publications, d'indiquer la qualité des rédacteurs, la date de publication et les sources des informations publiées.

Comme le montre la figure 1.2, les GBPC suivent un cycle de vie en plusieurs étapes :

1. En fonction des besoins exprimés et des directives nationales, le personnel d'Oncolor propose la création de nouveaux guides. Si la proposition est acceptée par les membres, un nouveau guide est créé.
2. La rédaction est réalisée de manière collaborative. Elle implique le personnel et les experts médicaux compétents réunis dans un groupe de travail dont les membres sont sélectionnés


REFERENTIELS
Tumeurs appendiculaires

Version téléchargeable et imprimable (PDF) | A propos de la version PDF (aide)

Ce référentiel, dont l'utilisation s'effectue sur le fondement des principes déontologiques d'exercice personnel de la médecine, a été élaboré par un groupe de travail pluridisciplinaire de professionnels des réseaux, lorrain ONCOLOR, alsacien CAROL et franc-comtois ONCOLIE, conformément aux données acquises de la science au ***10 octobre 2008***

- Généralités
- Anatomo-pathologie
- Bilan initial
- Critères RCP
- Classifications
- Traitements
- Surveillance
- Cancers métastatiques
- Bibliographie
- Rédacteurs

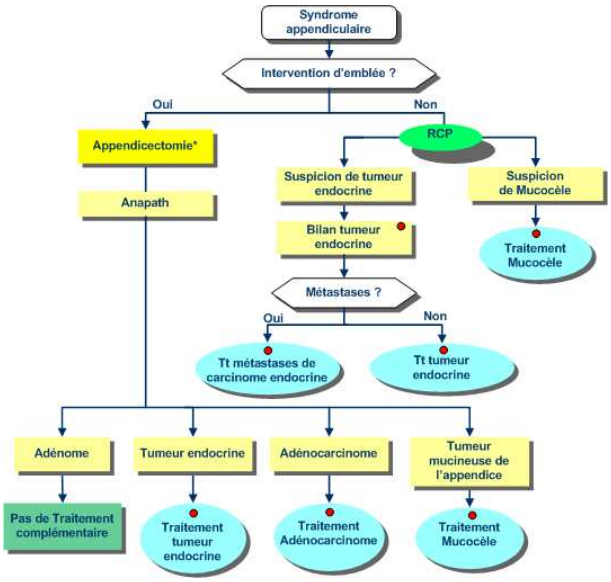
Cliquez-ici

Date de révision : ***10 octobre 2008***

Généralités

- Ce référentiel propose la conduite à tenir diagnostique et thérapeutique devant une tumeur de l'appendice.
- Ces tumeurs sont le plus souvent de découverte fortuite lors d'une appendicectomie pour syndrome appendiculaire ou lors d'une exploration chirurgicale de l'abdomen.

(a) Exemple d'introduction d'un GBPC.



(b) Exemple d'arbre de décision.

Classifications

- La classification TNM pour les adénocarcinomes de l'appendice est identique à celle utilisée pour les adénocarcinomes colorectaux (classification TNM 2002).
- Il n'y a pas de classification TNM pour les tumeurs endocrines.

Classification OMS

0	Capable d'une activité identique à celle précédent la maladie, sans aucune restriction
1	Activité physique diminuée mais ambulatoire et capable de mener un travail
2	Ambulatoire et capable de prendre soin de soi, incapable de travailler. Alité < 50 % de son temps
3	Capable de seulement quelques soins personnels. Alité ou chaise > 50 % du temps
4	Incapable de prendre soin de lui-même, alité ou chaise en permanence

(c) Exemple d'une classification.

FIGURE 1.1 – Extraits du GBPC sur les tumeurs appendiculaires, disponible dans ce format sur le site Web d'Oncolor avant avril 2012.

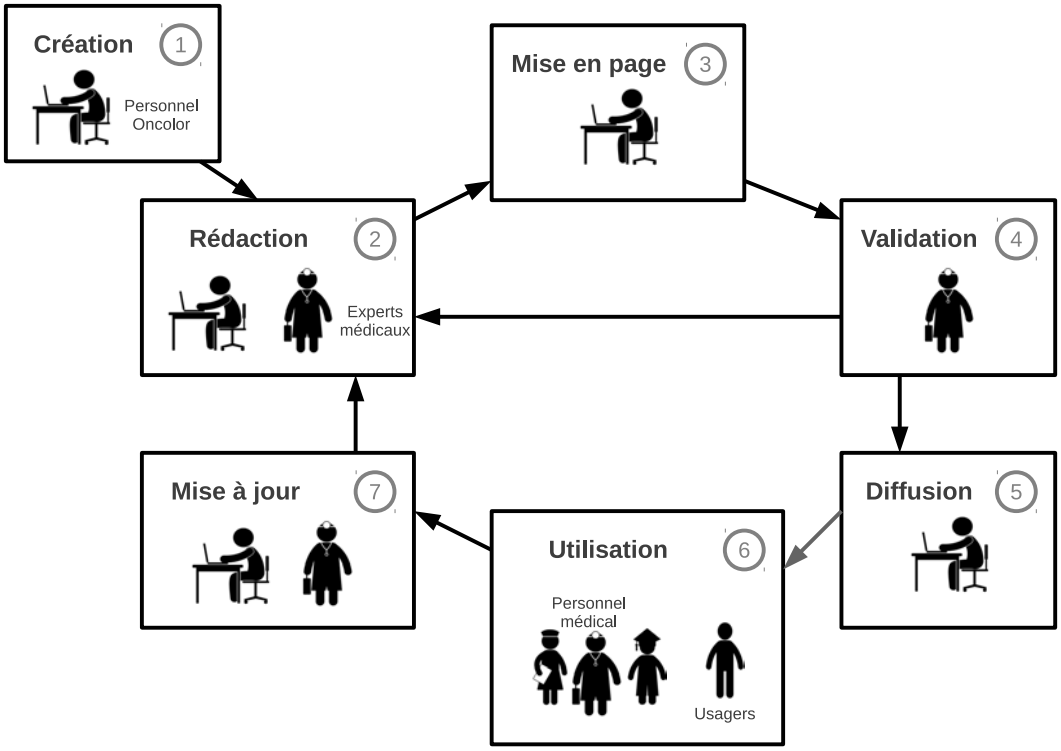


FIGURE 1.2 – Cycle de vie d'un GBPC édité par Oncolor.

par le coordinateur en charge du guide. Généralement, les médecins en santé publique proposent la structure du GBPC, qui est complétée par les experts. Les échanges sont coordonnés par le personnel d'Oncolor et se font généralement lors de réunions en petits comités, par l'intermédiaire de courriels, ou d'annotations sur des versions papier.

3. Un travail de mise en page est effectué au sein du réseau, de manière à fournir deux versions du guide : une version HTML et une version PDF imprimable. La qualité des PDF créés a notamment permis l'édition des guides sous forme de petits livres. L'édition est faite grâce à des outils de publication assistée par ordinateur (PAO) commerciaux : Visio (édité par Microsoft) pour les éléments graphiques, Dreamweaver (édité par Adobe) pour la version Web et Acrobat (édité par Adobe) pour la version PDF.
4. Les guides doivent ensuite être validés avant de permettre leur diffusion. Cette validation est généralement décidée lors de séminaires régionaux (ou inter-régionaux dans certains cas) par les experts médicaux. Si un consensus n'est pas obtenu parmi les experts, les guides sont renvoyés à l'étape de rédaction.
5. Lorsqu'ils sont validés, les GBPC sont autorisés à être diffusés en version papier et en version électronique sur l'Internet. Créé au milieu des années 90, le site d'Oncolor a été construit en HTML et, en suivant les évolutions du standard et des technologies, a intégré progressivement du CSS, du javascript, du XHTML et de l'ASP. Les GBPC y sont accessibles de manière ordonnée, selon une liste alphabétique, par des mots-clés ou en fonction d'une classification systématique.
6. Ainsi, les guides sont disponibles et librement utilisables par les professionnels de la médecine et les usagers. Lors de leur utilisation, les observations des membres d'Oncolor peuvent

mener au déclenchement d'une procédure de mise à jour.

7. Un processus de mise à jour peut être déclenché dans plusieurs cas : soit suite à des demandes d'utilisateurs, soit suite à une évolution du contexte local (comme par exemple la disponibilité d'un nouveau type de pratique ou de matériel), soit suite à une évolution majeure des données de la science remettant en cause les recommandations proposées, soit suite à l'échéance de la date de validité du référentiel. La mise à jour des GBPC suit alors la même procédure de rédaction et de validation que lors de sa création.

1.2.2.3 Vers l'utilisation d'outils collaboratifs

Avec le temps et le mélange des technologies, le site Web d'Oncolor est devenu de plus en plus difficile à maintenir. De plus, la coordination des groupes de travail est difficile à mener, par manque d'outils adéquats. Il n'existe pas non plus d'outils spécifiques permettant un suivi continu des GBPC, en conservant par exemple les remarques des praticiens en dehors des périodes de rédaction.

Dans ce contexte, il est apparu pour Oncolor le besoin d'un outil collaboratif pour servir de support à la création et à la mise à jour des GBPC. Cet outil doit prendre en charge l'édition collaborative des guides et de leurs composants, favoriser les échanges au sein du groupe de travail, permettre un suivi continu des guides et simplifier la maintenance technique des contenus en ligne. L'outil doit aussi respecter les exigences de transparence et de qualité de la certification HonCODE. Il convient donc de ne réserver l'accès qu'à des membres identifiés dont la compétence est certifiée. De plus, l'historique des modifications doit pouvoir être tracé.

1.2.3 Vers les guides de bonnes pratiques informatisés

L'évolution de l'informatique médicale tend désormais à favoriser l'informatisation de GBPC. Deux approches pour la formalisation des guides de bonnes pratiques informatisés (GBPI) sont distinguées par Georg [2006]. La première est centrée sur les connaissances, c'est-à-dire que son objectif est de proposer des méthodes et des outils pour l'édition directe de connaissances formalisées utilisables par des agents logiciels. Cette approche, qui ne prend pas en compte les GPBC pré-existants, fait l'objet du chapitre 4. La seconde est dite documentaire, c'est-à-dire qu'elle préserve le document comme support de présentation. Le but est d'enrichir les GBPC d'annotations en fonction des besoins identifiés des utilisateurs, tout en préservant leur forme originale de publication. Plusieurs travaux ont déjà exploité cette approche.

La plus ancienne approche documentaire est présentée par Hagerty *et al.* [2000] où les auteurs proposent d'étendre les GBPC initialement écrits en HTML à l'aide du langage HGML, un langage de balisage spécifique. Par défaut, les balises (par exemple, la balise <`recommandation`>) ne modifient pas l'aspect du document et permettent d'identifier la structure et les points les plus importants du document. Le but est de fournir de nouvelles interfaces pour le document en se servant des données contenues dans les annotations.

GEM[6] [Shiffman *et al.*, 2000] est l'implémentation la plus connue de l'approche documentaire. Fondée sur XML, elle propose d'intégrer le document à une hiérarchie le décrivant. Cette hiérarchie, présentée dans la figure 1.3, propose une structure pour les méta-informations du document (auteur, méthode de développement, etc.). En particulier, le composant *Knowledge Components* permet d'intégrer les recommandations. Pour celui-ci, une sous-hiérarchie est proposée

6. *Guideline Elements Model.*

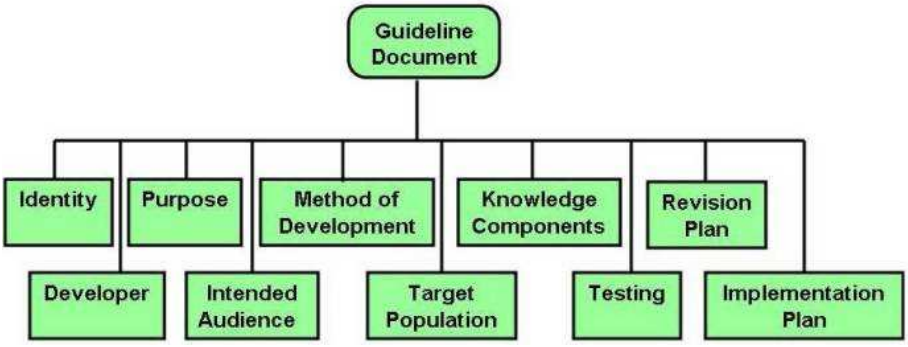


FIGURE 1.3 – Hiérarchie de document de GEM. Source : Tran *et al.* [2009].

pour décrire le processus de prise de décision, en déterminant des recommandations impératives ou conditionnelles.

Une autre approche, proposée par Georg et Jaulent [2005], utilise l'environnement G-DEE [7]. Dans ce cas, le but est d'extraire les connaissances du GBPC en utilisant des techniques de traitement automatique du langage. Utilisé dans le cadre de documents francophones, il permet la création semi-automatique de GBPI au format GEM.

1.3 Le Web sémantique

Support technologique de ce travail, le Web sémantique est un ensemble de technologies destiné à favoriser l'échange de connaissances en ligne. Après une présentation plus complète (section 1.3.1), sont exposés les principaux standards du Web sémantique (section 1.3.2) puis deux de ses déclinaisons, le Web de données (section 1.3.3) et le Web social et sémantique (section 1.3.4).

1.3.1 Présentation générale du Web sémantique

Dans la plupart des sites Web, les informations sont représentées en langage naturel ou par l'intermédiaire d'objets multimedia, destinées à être uniquement compréhensibles par un public humain. Le but du Web sémantique est de permettre aux machines de « comprendre » ces informations. Le Web sémantique est défini par Berners-Lee *et al.* [2001] comme « une extension du Web courant, dans laquelle on donne à une information un sens bien défini pour permettre aux ordinateurs et aux gens de travailler en coopération. » [8] En d'autres termes, on cherche à ajouter des significations compréhensibles par des machines aux informations du Web actuel de manière à ce que des ordinateurs puissent comprendre les documents présents sur le Web et ainsi accomplir des tâches qui n'auraient pu être réalisées que par un opérateur humain.

Cet objectif peut être illlustré par l'achat d'un trajet en ligne. Pour obtenir ses billets, un utilisateur doit d'abord se renseigner sur les modes de transport disponibles à son lieu de départ et de destination. Il doit ensuite consulter différents sites de vente de compagnies de transport pour pouvoir obtenir et comparer les différents tarifs, les horaires de déplacement et les préférences de

7. *Guidelines Document Engineering Environment.*

8. *« an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. ».*

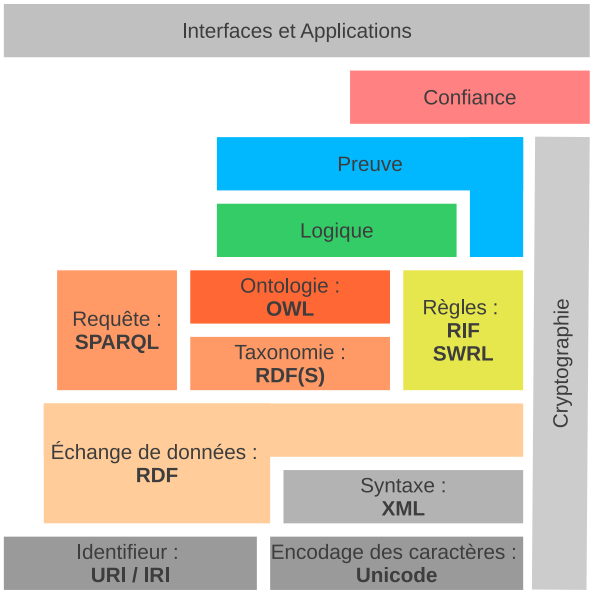


FIGURE 1.4 – *Layer cake* du Web sémantique, traduit et adapté de Semanticweb.org [2011].

confort. Il est facile d'imaginer des agents logiciels réalisant ces tâches répétitives. Toutefois, pour que ces agents puissent parvenir à les accomplir, il est nécessaire que les données utiles soient disponibles en ligne dans un format compréhensible par une machine. L'objectif du Web sémantique est de fournir ces formats, les outils mis à disposition et les méthodes de communication et de raisonnement qui permettront une telle interopérabilité entre les systèmes d'information.

1.3.2 Les standards du W3C

Pour atteindre l'objectif du Web sémantique, le W3C [9] [W3C, 2013] propose des standards pour la représentation, la visualisation et la diffusion des connaissances, ainsi que pour le raisonnement. Usuellement, ces standards sont représentés selon le célèbre *layer cake* de la figure 1.4.

Ce « gâteau » correspond à une illustration de la hiérarchie des langages du Web sémantique où chaque couche exploite les possibilités des couches plus basses. Il montre comment ces technologies se combinent pour atteindre les objectifs du Web sémantique énoncés précédemment. À l'origine proposée par Tim Berners-Lee, ce schéma évolue en fonction des avancées technologiques.

Encodage des caractères. Le but de cette couche est de fournir un encodage de caractères compatible avec un maximum de symboles issus des différents langages humains. La norme Unicode est proposée comme standard par le W3C.

Identifiant. Le but d'un identifiant est de fournir un nom unique à chaque ressource. Il repose sur les URI [10] [URI Planning Interest Group, 2001] qui sont des chaînes de caractères qui permettent de localiser une ressource sur un réseau, et les IRI [11] qui sont une internationalisation des précédentes en prenant en compte l'entièreté d'Unicode.

9. *World Wide Web consortium.*
10. *Uniform Resource Identifiers.*
11. *Internationalized Resource Identifiers.*

Syntaxe. Pour permettre les échanges de données, il faut définir la syntaxe de cette communication. Le format XML [Maler *et al.*, 2004] est la syntaxe d'échange la plus populaire sur l'Internet. L'utilisation des balises peut être contrainte par l'utilisation de DTD [12] ou de XML Schema, ce qui permet de définir une syntaxe propre à un domaine d'utilisation. Dans cette couche, aucune notion de sémantique n'est prise en compte.

Échange de données. Cette couche est la première qui soit propre au Web sémantique. Elle introduit la notion de sémantique en prenant en compte les relations entre les entités. Elle repose sur le *framework* RDF, présenté dans la section 1.3.2.1.

Taxonomie et ontologie. Pour stocker les connaissances nécessaires à ce type d'applications, le Web sémantique repose sur l'utilisation d'ontologies et de taxonomies. Une taxonomie est une liste de termes contrôlés organisés de façon hiérarchique permettant de créer des classifications dans un domaine. Selon la définition la plus communément admise en informatique, une ontologie est « une spécification explicite d'une conceptualisation » [13] [Gruber, 1993]. Une ontologie contient une description formelle des concepts d'un domaine, ainsi que les relations entre ceux-ci. De manière idéale, une ontologie doit être un modèle réutilisable et partagé des connaissances afin de servir de base commune à l'ensemble du domaine qu'elle représente. RDFS et OWL sont des langages permettant la création de telles bases de connaissances. Ils sont présentés respectivement dans les sections 1.3.2.1 et 1.3.2.2.

Requête. Pour interroger les connaissances, un langage compatible avec le *framework* RDF est nécessaire. SPARQL, présenté dans la section 1.3.2.3, est un tel langage.

Règles. RIF [Boley et Kifer, 2009] et SWRL [Horrocks *et al.*, 2004] permettent l'utilisation de règles. Ils permettent d'exprimer des restrictions sur les connaissances dont l'expression n'est pas possible dans les logiques de descriptions sur lesquelles repose OWL.

Logique et preuve. Un système de raisonnement utilisant les ontologies peut inférer de nouvelles connaissances. Ainsi, un agent logiciel utilisant un tel système peut déduire si une ressource particulière satisfait ses besoins. Le but de cette couche est de fournir les outils nécessaires à ce fonctionnement.

Cryptologie. Transversale aux couches précédentes, la cryptologie a pour but de veiller à l'intégrité des données consultées en ligne.

Confiance. Le but de cette avant-dernière couche est de fournir l'assurance de la qualité d'un service, c'est-à-dire de donner à un utilisateur les éléments lui permettant d'évaluer le niveau de confiance qu'il peut avoir en une information.

Interface et application. De nombreuses applications utilisant les techniques du Web sémantique sont déjà disponibles. Les wikis sémantiques, présentés dans le chapitre suivant, en forment un exemple.

12. *Document Type Definitions.*
13. « *an explicit specification of a conceptualisation* ».

1.3.2.1 Le *framework* RDF

RDF [14] est un modèle de représentation des méta-données pour les ressources du Web [Miller et Manola, 2004]. Dans cette section, nous présentons les éléments de RDF nécessaires à la compréhension de cette thèse.

Les triplets RDF. La construction de base de RDF est le triplet (*sujet, prédicat, objet*). Le sujet est la ressource que l'on souhaite décrire, représentée par son identifiant unique, c'est-à-dire une URI. Le prédicat, également présenté sous forme d'URI, définit la propriété qui s'applique à la ressource. Il établit une relation entre le sujet et l'objet en faisant de l'objet une caractéristique du sujet. Quant à l'objet, il peut prendre deux formes : soit une ressource présentée sous forme d'URI, soit un littéral, c'est-à-dire un objet typé, comme par exemple une chaîne de caractères ou un nombre entier.

Ainsi, si on considère la phrase « D'Artagnan est un héros des Trois Mousquetaires », elle pourra être représentée par le triplet :

```
(http://oeuvres.dumas.fr/ontologie.owl#D_ARTAGNAN,
 http://oeuvres.dumas.fr/ontologie.owl#estUnHerosDe,
 http://oeuvres.dumas.fr/ontologie.owl#LES_TROIS_MOUSQUETAIRES)
```

où
- « `http://oeuvres.dumas.fr/ontologie.owl#` » est l'espace de nom de l'ontologie des œuvres d'Alexandre Dumas,
- « `D_ARTAGNAN` » est la ressource représentant le personnage *d'Artagnan* et le sujet du triplet,
- « `estUnHerosDe` » est le prédicat signifiant « *est un héros de* »,
- « `LES_TROIS_MOUSQUETAIRES` » est la ressource représentant le roman *Les Trois Mousquetaires* et l'objet du triplet.

Notations. Il existe plusieurs notations pour RDF, comme l'illustre la figure 1.5. Deux sont des standards validés par le W3C : la syntaxe RDF/XML [Swick et Lassila, 1999], la plus commune, proposée dans le but de sérialiser les documents RDF en XML (voir l'exemple de la figure 1.5a) et la syntaxe N3 [Berners-Lee et Connolly, 2008] qui est conçue pour favoriser la lisibilité par les humains (voir l'exemple de la figure 1.5b).

Le vocabulaire RDFS. RDFS (*RDF Schema*) est un vocabulaire accompagné d'une sémantique formelle qui étend RDF en définissant un ensemble de classes et de propriétés. Ainsi, il permet notamment la création de hiérarchies de classes et de propriétés et fournit des définitions de littéraux, de domaines et de co-domaines des propriétés. Il permet également de définir le type d'une ressource et fournit quelques propriétés utiles, qui permettent par exemple d'ajouter une étiquette ou des commentaires à une ressource.

Si on applique le vocabulaire RDFS à la base de connaissances de la figure 1.5, on peut l'affiner en ajoutant de nouvelles connaissances. Ainsi, dans la figure 1.6, nous pouvons définir une classe `Personnage` et dire que `D_ARTAGNAN` en est un membre. De même, `LES_TROIS_MOUSQUETAIRES` est de type `Roman_Aventure` qui est une sous-classe de `Roman`. On peut également mieux définir la propriété `estUnHerosDe` en définissant son domaine, `Personnage`, et son co-domaine, `Roman`.

14. *Resource Description Framework*

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:od="http://oeuvres.dumas.fr/ontologie.owl#">
   <rdf:Description rdf:about="od:D_ARTAGNAN">
      <od:estUnHerosDe rdf:resource="od:LES_TROIS_MOUSQUETAIRES"/>
   </rdf:Description>
</rdf:RDF>
```

(a) Exemple de RDF avec la syntaxe RDF/XML.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix od: <http://oeuvres.dumas.fr/ontologie.owl#>.

<od:D_ARTAGNAN> od:estUnHerosDe <od:LES_TROIS_MOUSQUETAIRES>.
```

(b) Exemple de RDF avec la syntaxe N3.

FIGURE 1.5 – Exemples de syntaxes RDF, la syntaxe RDF/XML (a) et la syntaxe N3 (b).

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix od: <http://oeuvres.dumas.fr/ontologie.owl#>.

<od:Personnage> rdf:type <rdfs:Class>.
<od:D_ARTAGNAN> rdf:type <od:Personnage>.

<od:Roman_Aventure> rdf:type <rdfs:Class>.
<od:Roman_Aventure> rdfs:subClassOf <od:Roman>.
<od:LES_TROIS_MOUSQUETAIRES> rdf:type <od:Roman_Aventure>.

<od:estUnHerosDe> rdf:type <rdf:Property>.
<od:estUnHerosDe> rdfs:domain <od:Personnage>.
<od:estUnHerosDe> rdfs:range <od:Roman>.

<od:D_ARTAGNAN> od:estUnHerosDe <od:LES_TROIS_MOUSQUETAIRES>.
```

FIGURE 1.6 – Exemple d'une ontologie RDFS en syntaxe N3.

1.3.2.2 Les langages de représentation OWL

Développé comme une recommandation du W3C [Schreiber et Dean, 2004], OWL est un langage de représentation des connaissances conçu pour définir des ontologies. Difficilement lisible, il est destiné à être exploité par des processus automatiques, en particulier parce qu'il s'appuie le plus souvent sur la syntaxe RDF/XML. OWL étend le vocabulaire de RDFS en proposant des propriétés et des classes qui permettent d'exprimer des relations plus complexes. En OWL, l'élément de base est l'axiome. Un axiome utilise des classes, des propriétés et des instances pour définir de nouvelles classes et propriétés plus complexes. La table 1.1 présente les principaux apports d'OWL, ainsi que leurs représentations en logique de descriptions, dont les relations avec OWL sont présentées dans le paragraphe suivant.

La norme d'OWL 1.0 définit trois sous-langages selon une expressivité croissante :

- OWL Lite est la version la plus simple. Il a été pensé de manière à faciliter la reprise des thésaurus et ne propose donc qu'une faible expressivité.
- OWL DL propose une expressivité accrue du langage, tout en conservant la décidabilité des raisonnements. OWL DL contient tous les constructeurs OWL, mais comporte quelques restrictions d'utilisation.
- OWL Full correspond à l'ensemble d'OWL, incluant tous ses constructeurs et l'ensemble de RDFS. L'inférence en OWL Full est indécidable.

OWL peut être un support de raisonnement et plusieurs moteurs d'inférences compatibles existent. Les principaux sont présentés par Mishra et Kumar [2011]. Il est à noter qu'OWL utilise l'hypothèse de monde ouvert : c'est-à-dire que si un énoncé ne peut pas être démontré comme vrai, il n'est pas pour autant considéré comme faux.

La norme OWL 2. La deuxième version d'OWL [Krötzsch *et al.*, 2009] offre de nombreuses améliorations. D'après [Yu, 2011], elle peuvent être réparties en cinq catégories :

- simplification des constructeurs avec l'intégration de modèles de construction ;
- amélioration de l'expressivité avec de nouveaux constructeurs permettant par exemple le chaînage de propriétés ,
- support étendu des types de données comme celles dédiées au temps ;
- amélioration des possibilités d'annotation en permettant l'annotation de triplets ;
- des nouveaux profils de langages plus adaptés aux applications :
  - OWL 2 EL est conçu pour les grandes ontologies où la classification est l'application principale. En prenant en compte les restrictions d'OWL 2 EL, la complexité des algorithmes de raisonnement standard est au maximum polynomiale.
  - OWL 2 QL est optimisé pour les applications intégrant des ontologies et des bases de données classiques. Dans ces applications, les interrogations sur les instances sont généralement la tâche principale pour lesquelles OWL 2 QL garantit des complexités polynomiales.
  - OWL 2 RL permet l'implémentation de règles dans les ontologies en conservant une complexité algorithmique polynomiale.

Liens avec les logiques de descriptions. OWL est un formalisme inspiré des logiques de descriptions [Baader *et al.*, 2003] (LD), une famille de logiques qui sont un fragment décidable de la logique du premier ordre. Ce langage de représentation des connaissances s'appuie sur une représentation structurée et une sémantique formellement et précisément définie. Une représentation en LD est composée de plusieurs types d'entités : des concepts, des rôles et des individus

Classe	**Syntaxe OWL**	**Syntaxe en LD**
Concept universel	`owl:Thing`	$\top$
Concept impossible	`owl:Nothing`	$\bot$
Constructeur	**Syntaxe OWL**	**Syntaxe en LD**
Intersection	`owl:intersection C D`	$C \sqcap D$
Union	`owl:unionOf C D`	$C \sqcup D$
Complément	`owl:complementOf C`	$\neg C$
Énumération	`owl:oneOf a b ...`	$\{a, b, ...\}$
Quantificateur universel	`owl:Restriction(R owl:someValuesFrom C)`	$\exists R.C$
Quantificateur existentiel	`owl:Restriction(R owl:allValuesFrom C)`	$\forall R.C$
Restriction à une valeur	`owl:Restriction(R owl:hasValue a)`	$\ni R.a$
Restriction de cardinalité	`owl:Restriction(R owl:cardinality n)`	$= nR$
Restriction de cardinalité maximum	`owl:Restriction(R owl:maxCardinality n)`	$\geqslant nR$
Restriction de cardinalité minimum	`owl:Restriction(R owl:minCardinality n)`	$\leqslant nR$
Axiome	**Syntaxe OWL**	**Syntaxe en LD**
Classe équivalente	`owl:equivalentClass C D`	$C \equiv D$
Classe disjointe	`owl:disjointWith C D`	$C \sqcap D \sqsubseteq \bot$
Égalité d'instance	`owl:sameAs a b`	$a \equiv b$
Instance différente	`owl:differentFrom a b`	$a \sqsubseteq \neg b$
Sous-propriété	`owl:subPropertyOf` R_1 R_2	$R_1 \sqsubseteq R_2$
Propriété équivalente	`owl:equivalentProperty` R_1 R_2	$R_1 \equiv R_2$
Propriété inverse	`owl:inverseOf` R_1 R_2	$R_1 \equiv R_2^-$
Transitivité	`owl:TransitiveProperty R`	$R^+ \sqsubseteq R$
Être une fonction partielle	`owl:FunctionalProperty R`	$\top \sqsubseteq \leq 1R$
Être l'inverse d'une fonction partielle	`owl:InverseFunctionalProperty R`	$\top \sqsubseteq \leq 1R^-$

TABLE 1.1 – Ressources introduites par OWL et leur syntaxe en logique de descriptions (LD). C et D sont des classes OWL (des concepts en LD), R, R_1 et R_2 sont des propriétés (des rôles en LD), a et b sont des instances (des individus en LD), n est un entier positif.

qui correspondent respectivement aux classes, propriétés et instances d'OWL. Ces éléments sont combinés pour produire des bases de connaissances. Celles-ci sont constituées de deux parties :

- une « *Terminological Box* » (TBox), où sont définis les axiomes terminologiques qui décrivent les connaissances générales d'un domaine ;
- une « *Assertional Box* » (ABox), qui contient les assertions, c'est-à-dire les connaissances factuelles du domaine.

La sémantique associée à une base de connaissances en LD est définie par des interprétations. Formellement, une interprétation est un couple $\mathcal{I} = (\Delta_\mathcal{I}, \cdot^\mathcal{I})$ où $\Delta_\mathcal{I}$ est un ensemble non vide et où $\cdot^\mathcal{I}$ est une fonction d'interprétation. La fonction d'interprétation $\cdot^\mathcal{I}$ fait correspondre à un concept un sous-ensemble de $\Delta_\mathcal{I}$, à un rôle une relation binaire sur $\Delta_\mathcal{I} \times \Delta_\mathcal{I}$ et à un individu un élément de $\Delta_\mathcal{I}$. On dit qu'une interprétation $\mathcal{I}$ est un modèle d'une base de connaissances BC si elle satisfait tous les axiomes et toutes les assertions de BC.

Des mécanismes de raisonnement peuvent être appliqués aux bases de connaissances. Des inférences peuvent être faites tant au niveau de la TBox que de la ABox. Les raisonnements en LD se fondent sur deux tests qui peuvent être combinés pour former des tests plus complexes :

- les tests de subsomption, noté $\models_\theta C \sqsubseteq D$, où θ est une ontologie, qui consistent à vérifier qu'un concept D subsume un concept C, c'est-à-dire que chaque individu de C est aussi un individu de D,
- les tests de satisfiabilité, qui vérifient la satisfiabilité d'un concept.

```
PREFIX od: <http://oeuvres.dumas.fr/ontologie.owl#>

SELECT ?personnages
WHERE
    { ?personnages od:estUnHerosDe od:LES_TROIS_MOUSQUETAIRES }
```

FIGURE 1.7 – Exemple de requête SPARQL.

1.3.2.3 Le langage d'interrogation SPARQL

SPARQL [15] est un langage d'interrogation pour RDF [Prud'hommeaux et Seaborne, 2008b]. Il permet d'interroger les sources de données RDF disposant d'un *SPARQL endpoint*, telles que les *RDFStores*. De manière schématique, on pourrait dire que SPARQL est pour les bases de connaissances ce que SQL est pour les bases de données. Sa syntaxe est d'ailleurs proche de SQL, disposant des clauses *SELECT* et *WHERE*. La figure 1.7 donne l'exemple d'une requête en SPARQL permettant de savoir quels sont les héros du roman *Les Trois Mousquetaires*.

SPARQL utilise également l'hypothèse de monde ouvert. Ainsi, pour la requête de la figure 1.7, la réponse ne contiendra que les héros connus du système.

1.3.3 Vers le Web de données

Le développement du Web sémantique au cours des années 2000 a conduit à une prolifération des ontologies, formant des ensembles de connaissances disparates et séparées. Cette évolution a conduit à l'apparition d'un Web sémantique séparé du Web classique, le premier étant destiné principalement aux échanges entre machines et le second aux échanges entre humains. Pour briser cette frontière, la première évolution proposée par le W3C est RDFa [16] [Birbeck et Adida, 2008]). RDFa étend XHTML [17] en permettant l'intégration de triplets RDF directement dans les pages Web. Le code RDFa, invisible pour l'utilisateur, peut faire l'objet d'un traitement automatique. Il est destiné à recueillir des méta-informations à propos de la page.

Berners-Lee [2006] a proposé une nouvelle manière de publier les données structurées sur le Web, celle du Web de données [18]. Dans ce cadre, l'Internet est vue comme une base de données globale où les données sont reliées à travers des liens hypertextes ayant une sémantique donnée. En se servant des mécanismes de HTTP, on peut proposer pour chaque objet ayant une URI une présentation à destination des humains et une page relayant les informations en RDF à l'intention des machines.

Sur ce principe, de nombreuses bases de connaissances ont émergé. Ces bases sont reliées entre elles autour du projet de base généraliste DBpedia [DBpedia, 2013], formant ainsi un nuage de connaissances comme l'illustre la figure 1.8.

15. SPARQL est également le nom du protocole conçu pour transmettre les requêtes [Feigenbaum et Johnston, 2009].
16. *Resource Description Framework attributes.*
17. XHTML est une extension de HTML fondée sur la syntaxe XML [Epperson *et al.*, 2006].
18. en anglais, *Linked Data.*

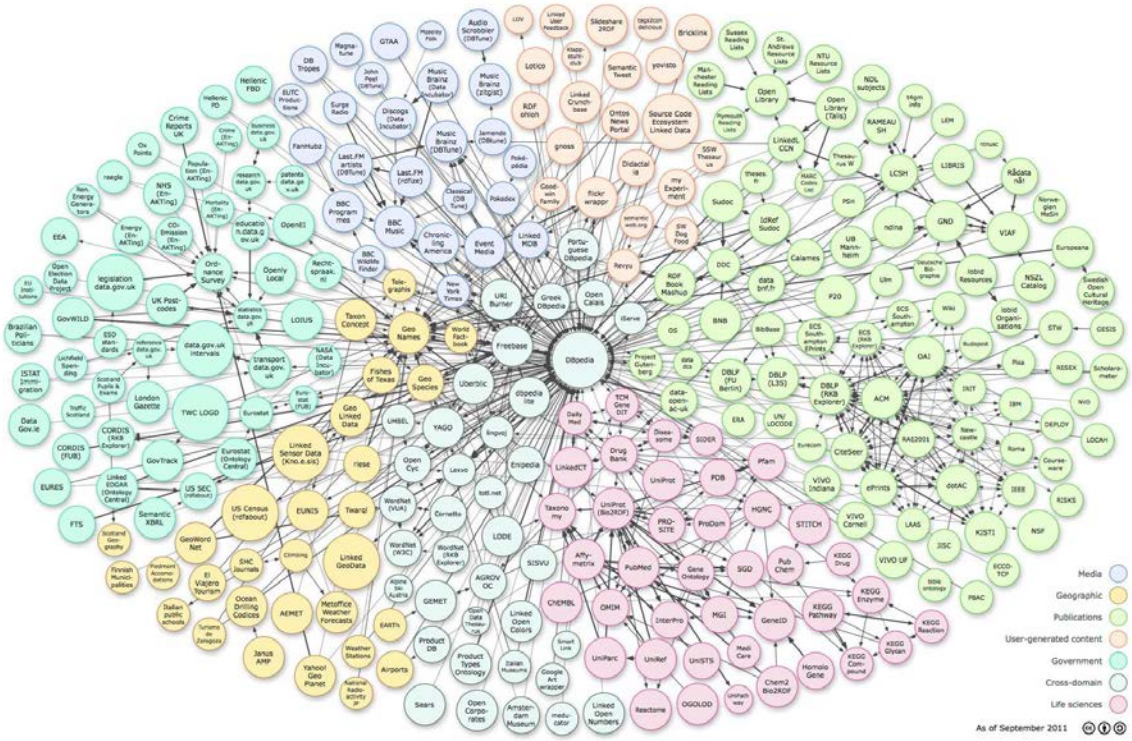


FIGURE 1.8 – Nuage du Web de données (source : `http://richard.cyganiak.de/2007/10/lod/`, 2011).

1.3.4 Le Web social et sémantique

De manière théorique, le Web social est défini comme l'ensemble des relations qui relient les personnes sur le Web [Appelquist *et al.*, 2010]. D'un point de vue plus pratique, le Web Social peut être considéré comme la vision d'un Web centré sur l'échange entre utilisateurs, par opposition à la vision d'un Web documentaire. Ainsi, le Web devient une plate-forme collaborative dont la valeur ajoutée est l'agrégation des contributions des utilisateurs. Les exemples d'application du Web social sont nombreux, comme les populaires Facebook, Twitter et Wikipédia. Techniquement, le Web social exploite les technologies et les usages du Web 2.0 [19] qui contribuent à améliorer les échanges sur l'Internet.

Une vision du Web social et sémantique est donnée par Gruber [2008] où l'intelligence collective du Web social est exploitée avec les technologies du Web sémantique. Le rôle du Web sémantique est de fournir les techniques pour capturer, stocker, distribuer et favoriser la communication des connaissances. Ainsi, ces technologies doivent « ajouter des données structurées relatives aux contenus des contributions des utilisateurs d'une manière qui permette un puissant traitement informatique » [20]. Cette vision du Web peut être illustrée par le projet DBpedia, qui est un exemple typique d'application du Web social et sémantique. Le but de ce projet est d'extraire des informations structurées construites collaborativement dans Wikipédia et de les rendre disponibles selon les principes du Web de données [Bizer *et al.*, 2009].

19. Voir [O'Reilly, 2005].
20. « *add structured data related to the content of the user contributions in form that enables more powerful computation.* » [Gruber, 2008].

1.4 Outils du Web sémantique pour l'informatique médicale

Le partage des informations est au centre des préoccupations de l'informatique médicale. Ces dernières années, l'intégration des techniques du Web sémantique a permis de nombreuses avancées, notamment pour l'encodage, le stockage et la réutilisation des connaissances médicales. Le but de cette section est de dresser un bref panorama de ces avancées, en présentant en particulier les thésaurus et ontologies médicales (section 1.4.1) et les applications en ligne qui leur sont associées (section 1.4.2), ainsi que les outils proposés dans le cadre du projet KASIMIR (section 1.4.3).

1.4.1 Thésaurus et ontologies médicales

La première création de classification médicale remonte à plus d'un siècle, avec la rédaction dès 1893 de la Classification Internationale des Maladies (CIM)[21], ce qui traduit le besoin d'établir un vocabulaire unique afin de normaliser l'utilisation des termes en médecine. Avec l'informatisation de la médecine, de nombreuses ressources sont ainsi apparues, généralement liées à des applications spécialisées. Au cours des dernières années, le Web sémantique a fourni un cadre technique à leur élaboration en proposant des formats permettant une expressivité en rapport avec les capacités d'analyse des machines.

1.4.1.1 Rôles en informatique médicale

Les thésaurus et les ressources ontologiques sont devenus des outils importants en informatique médicale, utilisés aussi bien dans le cadre des soins au patient que dans celui de la recherche biomédicale. Bodenreider [2008] décrit les rôles que peuvent jouer ces ressources dans les systèmes médicaux. Pour plus de clarté, ces rôles sont divisés en trois domaines d'application : le premier est la gestion des connaissances, le deuxième est l'interopérabilité sémantique, et le troisième comprend l'aide à la décision et le raisonnement. Par la suite, pour chacun de ces domaines, le rôle est expliqué et des exemples d'application sont donnés.

Rôle dans la gestion des connaissances. Les ontologies médicales sont utilisées pour fournir le vocabulaire du domaine, c'est-à-dire une liste de noms d'entités dont la sémantique est définie. Ainsi, elles servent à l'accès, l'annotation et l'alignement des ressources. La section 1.4.2 évoque des exemples de gestion de ressources bibliographiques optimisées par l'emploi de thésaurus ou d'ontologies.

Rôle dans l'interopérabilité sémantique. Les systèmes d'information en médecine ont besoin d'échanger des données structurées mais l'intégration des données n'est possible qui si elles sont standardisées. Les ontologies médicales peuvent fournir cette structure, comme par exemple pour des dossiers patient informatisés. Dans ce cas, le but est de conserver l'ensemble des informations concernant la santé du patient d'une manière formelle. Ce dossier a donc vocation à être échangé entre les différents systèmes d'information utilisés par les praticiens. Les informations utiles à la description de la santé du patient sont encodées en fonction de ressources terminologiques utilisables par les différents systèmes.

21. en anglais : *International Statistical Classification of Diseases and Related Health Problems* (ICD).

Rôle dans l'aide à la décision et le raisonnement. Les ontologies médicales stockent les connaissances du domaine et leurs définitions formelles, ce qui peut favoriser la découverte de nouvelles connaissances grâce à des processus de raisonnement et ainsi fournir de nouveaux services. Les mécanismes de raisonnement peuvent par exemple être utilisés pour la sélection et l'agrégation des données, dans le cadre de la sélection d'un groupe de tests cliniques et de la définition des caractéristiques de ce groupe. Une autre application est liée aux GBPI : une ontologie peut fournir les outils pour encoder une situation médicale et contenir des règles déterminant l'exécution d'alertes à destination des praticiens. Ainsi, l'utilisation d'ontologies médicales ouvre la voie aux techniques de l'intelligence artificielle. Par exemple, lorsqu'elles sont associées aux techniques de traitement automatique du langage, il est possible de créer des systèmes médicaux de questions réponses et des outils de résumé automatique pour la littérature médicale. Un exemple de découverte de connaissance est évoqué par Bodenreider [2008] où des ontologies médicales sont utilisées dans la recherche biomédicale pour identifier des associations entre phénotype et génotype.

1.4.1.2 Conception des ontologies médicales

Comme souligné par Freitas *et al.* [2009], une distinction stricte entre thésaurus et ontologie en médecine n'est pas toujours applicable. En effet, les ressources majeures en informatique sont généralement des vocabulaires initialement conçus comme des thésaurus. En fonction des besoins des utilisateurs et de l'arrivée des technologies permettant la gestion de la sémantique, elles ont évolué vers des ontologies, intégrant au fur et à mesure plus d'expressivité tout en s'efforçant de conserver une rétro-compatibilité.

Quatre types d'ontologies utilisées dans les systèmes d'information médicaux sont identifiées par Bodenreider et Burgun [2005] :

- Les ontologies générales, qui ne sont pas forcément conçues pour les applications de santé, sont indépendantes des tâches. Elles fournissent les concepts généraux utiles, comme par exemple des concepts relatifs au temps ou à l'espace. Ces concepts sont appelés à être exploités par des ontologies médicales auxquelles elles sont reliées.
- Les ontologies du domaine concentrent toutes les connaissances sur un domaine spécifique, comme par exemple la cancérologie ou la génétique.
- Les ontologies d'application sont spécifiques à un système ou à une application, et particulièrement à une tâche pour laquelle elles ont été créées.
- Les ontologies de référence sont des petits modules, indépendants du contexte, réutilisables dans différents domaines.

Les ontologies médicales peuvent également être caractérisées selon le niveau de granularité, souvent dépendant de leur type, de l'application et du domaine pour lequel elle a été conçue. En effet, si on souhaite par exemple modéliser le corps humain, sa description au niveau anatomique et sa description au niveau cellulaire se feront à des niveaux de granularité différents. Ces niveaux seront difficiles à joindre si on souhaite utiliser les deux modélisations au sein d'une même application [Fung et Bodenreider, 2012].

Les deux principales difficultés dans l'utilisation des ontologies médicales sont identifiées par Fung et Bodenreider [2012]. La première est la prolifération des ontologies médicales. Plusieurs centaines sont disponibles, sans qu'il existe de métrique communément admise pour leur validation et leur certification [Bodenreider et Stevens, 2006]. La deuxième difficulté est le manque d'interopérabilité entre ces ontologies. Ainsi, Bodenreider et Burgun [2005] montrent que l'alignement du concept « *blood* » entre différentes ontologies de l'UMLS [22] mène à des incohérences.

22. UMLS est décrit dans la section suivante.

1.4.1.3 Exemples de RTO largement utilisées

De nombreux thésaurus et ontologies médicales ont déjà été créés, dont beaucoup sont présentés dans les états de l'art proposés par Baneyx [2007] et plus récemment par Freitas *et al.* [2009]. La table 1.2, qui présente les ressources les plus utilisées, montre la variété des ressources à disposition et leur dépendance au contexte de développement.

FMA. L'ontologie FMA [23] [UW Structural Informatics group, 2013] est un modèle didactique sur la structure macroscopique du corps humain. Les concepts sont organisés selon deux hiérarchies, une représentant une classification anatomique où la relation de sous-classe est de type « sous-classe de » et une autre hiérarchie ordonnée selon une relation « partie de » [24]. D'autres relations sémantiques (e.g. « adjacent à » [25]) et attributs complètent les descriptions. Alors que des expériences de raisonnement ont été validées, l'exploitation de FMA peut toutefois mener à des inconsistances [Freitas *et al.*, 2009].

openGalen. openGalen [26] [OpenGalen, 2013] est un projet européen d'ontologie médicale représentée en LD. Son but est de représenter les concepts indépendamment des langues pour améliorer l'interopérabilité, dans le cadre notamment du dossier patient et des systèmes d'aide à la décision. Son modèle peut être divisé en trois parties : une ontologie de haut niveau qui fournit les connaissances de catégorisation, le *GALEN Common Reference Model* qui contient des définitions de concepts communs à toutes les disciplines et des extensions spécifiques aux sous-domaines telles que la chirurgie ou l'anatomie.

Gene Ontology. Gene Ontology [Cimino et Zhu, 2006; Gene Ontology, 2013] (GO) est une terminologie destinée à standardiser la représentation des gènes et des produits géniques de toutes les espèces. À l'image de FMA, les concepts sont organisés selon une double hiérarchie ordonnée avec les relations « sous-classe de » et « partie de ».

CIM-10. CIM-10 [27] [WHO, 2013] est la dixième révision de la CIM, la plus ancienne classification médicale. Elle permet le codage des maladies, des traumatismes et de l'ensemble des motifs de recours aux services de santé. La CIM-10 est organisée selon une hiérarchie monoaxiale de type « sous-classe de » dont le premier niveau est composé de 22 chapitres (e.g. « Tumeurs », « Maladies de l'appareil digestif »).

MeSH. Le MeSH [28] [NCBI, 2013c] est un thésaurus conçu pour l'indexation de publications médicales et la recherche d'informations pour les sciences de la vie. Les concepts y sont classés de manière hiérarchique selon 16 axes. La hiérarchie de concepts est organisée de manière à ce que tous les documents indexés par un concept soient également pertinents pour tous ses concepts parents. Les concepts sont décrits par des définitions textuelles, des attributs (exprimant par exemple des relations de synonymie) et des « *qualifiers* » permettant de préciser le contexte (e.g. « épidémiologie »).

23. *Foundational Model of Anatomy.*
24. Les noms des relations ont été traduits. Les noms originaux de « sous-classe de » et « partie de » sont respectivement « *is-a* » et « *part-of* ».
25. Le nom original est « *adjacent_to* ».
26. Galen pour *Generalized Architecture for Languages Encyclopaedias and Nomenclature.*
27. en anglais *International Statistical Classification of Diseases and Related Health Problems* (ICD-10).
28. *Medical Subject Headings.*

NCI Thesaurus. Le NCI Thesaurus [NCI, 2013] est une ontologie utilisée pour indexer les informations concernant les essais cliniques en cancérologie. Son organisation reprend les structures classiques des ontologies (classes, propriétés, contraintes) et offre une couverture large du domaine en décrivant les maladies, les médicaments, les thérapies, etc. Son expressivité est limitée, puisqu'elle utilise OWL-Lite. Toutefois, une étude a mis en avant de nombreuses erreurs et inconsistances dans l'ontologie [Ceusters *et al.*, 2005].

SNOMED CT. SNOMED CT [29] [IHTSDO, 2013] est une ontologie fondée sur les LD. Elle est destinée à être utilisée dans le contexte des soins de santé cliniques, en particulier pour l'encodage des dossiers patient. Ses concepts sont organisés selon 16 axes représentant une hiérarchie de type « sous-classe de ». Décrits par des définitions syntaxiques et des attributs additionnels, les concepts peuvent être combinés pour exprimer de nouveaux concepts, et reliés entre eux selon 50 types de liens prédéfinis.

UMLS. UMLS [30] [NLM, 2013b] est un metathésaurus dans le sens où son but est d'unifier plusieurs terminologies (dont MeSH, SNOMED, etc.) en un seul espace. Pour ce faire, il intègre les différents concepts et les relations dans un thésaurus unique, en fusionnant les concepts semblables et en les reliant à travers un réseau sémantique reposant sur des relations très générales (e.g. « body part »). Toutefois, étant donné la taille et la complexité du thésaurus résultant, de nombreuses erreurs peuvent être détectées, ainsi que des ambiguïtés, des redondances et des relations hiérarchiques cycliques.

1.4.2 Applications de l'informatique médicale en ligne

Les ressources ontologiques sont déjà largement utilisées en informatique médicale. Dans cette section, nous nous intéressons spécialement aux applications disponibles en ligne s'inscrivant dans le contexte du Web sémantique. Deux types d'applications sont présentées :

- les applications favorisant l'utilisation des ontologies médicales, en proposant des dépôts et généralement des services de consultation et d'interrogation automatique,
- les applications exploitant des ontologies médicales, en particulier les ressources bibliographiques réutilisables dans le reste de ce travail.

1.4.2.1 Dépôts d'ontologies médicales

OBO Foundry. OBO Foundry [31] [OBO Foundry, 2013] est une plate-forme en ligne, présentée comme une librairie pour les ontologies biomédicales. Elle regroupe plusieurs ontologies (dont FMA et GO) développées selon les bonnes pratiques du *framework* OBO, édictant des principes notamment sur les définitions des entités et la documentation des ontologies. De plus, toute ontologie souhaitant être hébergée doit prendre en compte les ontologies déjà présentes en réutilisant les concepts et les propriétés précédemment définis. Son but est d'améliorer l'interopérabilité des systèmes médicaux en proposant des ontologies compatibles les unes avec les autres. Plus de cinquante ontologies sont disponibles en ligne, chacune représentant une spécialité ou une application médicale.

29. *Systematized Nomenclature of Medicine - Clinical Terms.*
30. *Unified Medical Language System.*
31. *Open Biomedical Ontologies Foundry.*

Ressource	Domaine d'application	Langues	Nombre de concepts	Formalisme	Hiérarchie de subsomption	Liens typés	Organisme
FMA	Anatomie	Multi-langue dont français	Plus de 75 000	OWL	Oui	Oui	Structural Information Group, University of Washington
openGALEN	Anatomie, chirurgie, maladie et soins	Multi-langue dont français	Plus de 10 000	GRAIL	Oui	Oui	GALEN Project
Gene Ontology	Génétique	Anglais	22 546	OBO/OWL	Oui	Non	The GO Consortium
CIM-10	Statistiques de santé, épidémiologie	42 dont français	Plus de 14 400	Classification	Non	Non	Organisation Mondiale de la Santé
MeSH	Indexation de la littérature médicale	Multi-langue dont français	24 767	Réseaux sémantiques	Non	Non	US National Library of Medicine
NCI Thesaurus	Cancérologie	Anglais	58 868	Logique de descriptions	Oui	Oui	US National Cancer Institute
SNOMED CT	Médecine clinique	Multi-langue dont français	310 314	LD	Oui	Oui	International Health Terminology Standards Development Organisation
UMLS	Intégration de terminologies biomédicales	Multi-langue dont français	1,4 Millions	Réseaux sémantiques	Oui	Oui	US National Library of Medicine

TABLE 1.2 – Caractéristiques de quelques unes des principales ressources ontologiques médicales. Sources : Bodenreider [2008], Freitas *et al.* [2009]

Bioportal. BioPortal [BioPortal, 2013; Noy *et al.*, 2009] est un espace de dépôt d'ontologies biomédicales. En plus de l'hébergement d'ontologies, BioPortal fournit de nombreux outils destinés à améliorer leur interopérabilité avec les systèmes d'information médicaux. En août 2012, il contenait 309 ontologies, dont quelques unes des plus connues comme MeSH et SNOMED, rassemblant plus de 5 millions de termes. Le portail fournit aux utilisateurs des outils pour rechercher, visualiser et naviguer dans les ontologies et propose des outils du Web social permettant de commenter, annoter, proposer des modifications. Les techniques du Web social sont également utilisées pour éditer des alignements entre ontologies. Enfin, BioPortal fournit une URI fixe pour chaque concept, ce qui les rend disponibles pour les applications du Web de données.

Portail Terminologique de Santé. Le Portail Terminologique de Santé [CiSMeF, 2013; Grosjean *et al.*, 2011] (PTS) est un serveur multi-terminologique francophone dont le but est de faciliter la navigation dynamique dans les ressources terminologiques et ontologiques médicales. Il exploite un grand nombre de ces ressources dont CIM-10, SNOMED, MeSH et FMA en les intégrant à des outils de recherche de concepts. Techniquement, le PTS exploite les concepts d'UMLS pour réaliser des alignements entre les ressources. Les ressources sont également reliées par un meta-modèle OWL les incluant toutes, et chacune est représentée par son propre modèle OWL. D'après Grosjean *et al.* [2011], le PTS est principalement utilisé dans deux contextes : pour l'apprentissage de la manipulation des terminologies médicales par des étudiants et pour l'aide à l'indexation de documents médicaux.

Les bases ouvertes du Web de données. L'évolution vers le Web de données a eu pour conséquence la sémantisation et l'ouverture de bases de données biomédicales permettant la mise à disposition de grandes bases RDF. Un exemple de processus d'ouverture d'une base biomédicale est présenté par Belleau *et al.* [2008]. De nombreuses bases sont ainsi disponibles pour la création de *mashup*[32], comme par exemple LinkedCT, qui indexe plus de 60 000 essais cliniques, et Drugbank, la base d'informations sur les médicaments qui contient plus de 1 150 000 triplets RDF. Plusieurs études présentent des utilisations de ces bases ouvertes, comme celle de Sonntag *et al.* [2012] où des données liées servent à l'annotation d'images en radioscopie.

1.4.2.2 Sites de références bibliographiques

Plusieurs sites Web indexent les ressources bibliographiques médicales. Ces ressources constituent les sources à partir desquelles sont contruits les RPC. Le plus célèbre est PubMed [NCBI, 2013b]. Il propose un moteur de recherche pour la base MEDLINE [NLM, 2013a], rassemblant des citations, des résumés et des annotations sur les articles internationaux de recherche biomédicale. Plus particulièrement, le moteur permet la construction de requêtes avancées équivalentes en expressivité à la logique propositionnelle. L'indexation de MEDLINE utilise le thésaurus MeSH, présenté précédemment.

À l'échelle nationale, CISMeF [Darmoni *et al.*, 2000] propose un service du même type à destination des professionnels de santé. Il indexe pour sa part les ressources médicales francophones disponibles en ligne. Pour ce faire, chaque site Web est indexé selon une notice descriptive reposant principalement sur MeSH. De plus, cette indexation tend à devenir multi-terminologique grâce à l'utilisation du PTS. Il dispose de son propre moteur de recherche, proposant des fonctionnalités analogues à celles du moteur de PubMed.

32. Un *mashup* est une application dont le contenu est issu de plusieurs sources de données.

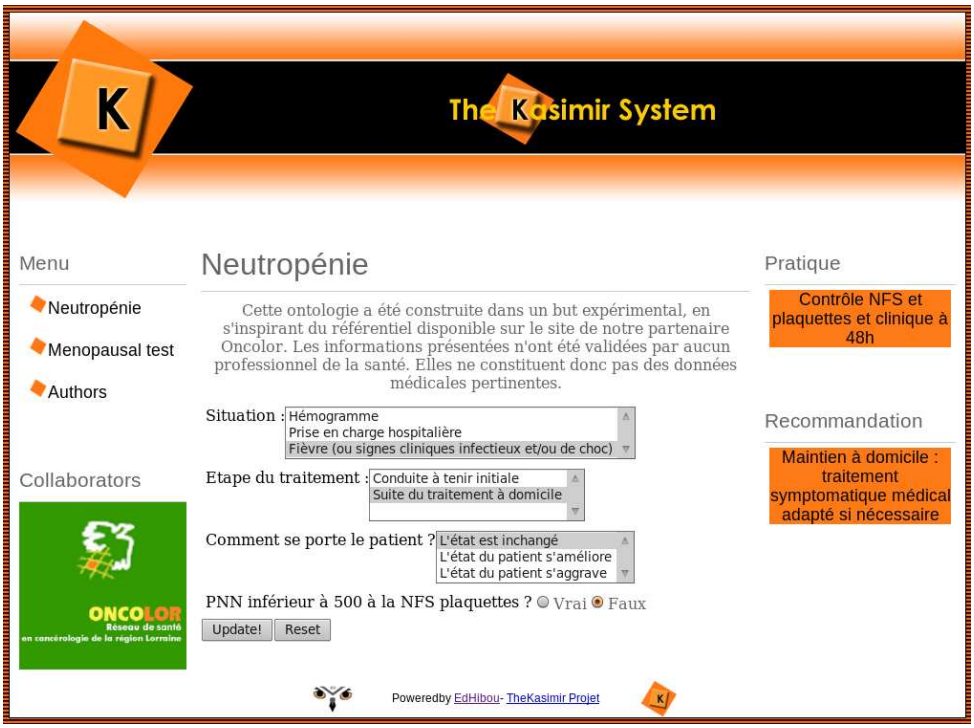


FIGURE 1.9 – Application d'EDHIBOU dans la cadre de la formalisation d'un GBPC sur la neutropénie.

1.4.3 Les réalisations du projet KASIMIR

Plusieurs outils ont déjà été développés dans le cadre du portail sémantique en cancérologie du projet KASIMIR, dont plusieurs sont décrits par d'Aquin [2005].

Parmi ces outils, KOWL est un serveur de connaissances pour le Web sémantique. Le temps d'une session, il permet le peuplement d'une ontologie OWL par l'ajout d'instances et d'assertions. Il permet également l'interrogation de cette ontologie par l'utilisation du langage de requête SPARQL. KOWL supporte également le raisonnement en prenant en charge la communication avec un moteur d'inférences. Dans le cadre de ce travail, le moteur d'inférences utilisé est Pellet [Sirin *et al.*, 2007b].

Le *framework* EDHIBOU [Badra *et al.*, 2008b] est un autre outil du projet KASIMIR, pouvant être implémenté en tant que service Web. Son but est de fournir une interface utilisateur pour KASIMIR qui permette la description d'une situation médicale et fournisse la recommandation adaptée. Techniquement, le principe de l'application est de créer une instance dans une ontologie OWL et de considérer les propriétés liées comme des questions auxquelles l'utilisateur peut répondre dans l'interface. EDHIBOU propose des vues annexes paramétrables afin de suivre l'évolution d'éléments dans l'ontologie en fonction des choix effectués par l'utilisateur, par exemple les classes d'une instance. Visuellement, les éléments graphiques contenus dans EDHIBOU sont personnalisables grâce à l'utilisation d'une ontologie dédiée dans laquelle les vues disponibles sont paramétrées et les composants graphiques de formulaires peuvent être choisis et configurés. De plus, l'emploi de feuilles de style CSS permet d'adapter l'aspect à toutes les applications Web. Pour effectuer ces tâches et construire l'interface en fonction des désidératas de l'utilisateur, le *framework* repose sur les capacités d'édition et d'inférences du serveur de connaissances KOWL.

Un exemple d'utilisation d'EdHibou est présenté dans la figure 1.9.

Toutefois, le projet ne propose pas d'outil pour l'acquisition des connaissances et aucun de ses travaux n'intègre de lien avec des RTO reconnues d'informatique médicale. En conséquence, l'utilisation des composants créés ne se limite qu'à quelques exemples de démonstration.

1.5 Conclusion partielle

Le projet Kasimir s'inscrit dans le cadre de la gestion des connaissances en oncologie. Plus précisément, son but est de fournir des outils et des méthodes pour la formalisation, le stockage et l'exploitation des connaissances décisionnelles. Afin d'y parvenir, de récents travaux ont mis en avant le besoin d'un portail sémantique. Le cadre technique de ce portail est fourni par le Web sémantique, un ensemble de standards permettant la formalisation et l'exploitation automatique des connaissances. La représentation des connaissances est un thème récurrent de l'informatique médicale, où bon nombre de thésaurus et d'ontologies ont déjà été créés. Dans le chapitre suivant, une solution originale du Web social et sémantique est présentée : les wikis sémantiques. Ces outils ayant atteint un bon stade de maturité, la suite de ce travail consistera à montrer de quelle manière ils peuvent être utilisés dans le cadre du projet Kasimir.

Chapitre 2

État de l'art des wikis sémantiques

Sommaire

Après presque 20 ans d'existence, les wikis sont désormais des systèmes reconnus dans le monde de l'Internet. En parallèle, le développement du Web sémantique depuis le début des années 2000 a ouvert de nouvelles perspectives. Ainsi sont nés les wikis sémantiques, dont la particularité consiste à formaliser le contenu des articles, notamment en caractérisant les relations entre ceux-ci. De nombreux moteurs de wiki sémantique ont vu le jour depuis 2003, avec plus ou moins de suivi et de succès.

Dans ce chapitre, nous présentons et comparons plusieurs de ces systèmes. La section 2.1 présente les wikis et évoque leurs forces et leurs limites. Dans la section 2.2, nous définissons la notion de wiki sémantique ainsi que les deux approches guidant leur conception, *Wikis for Ontologies* et *Ontologies for Wikis*. Enfin, la section 2.3 présente les principaux moteurs de wikis sémantiques et donne des éléments de comparaison. Cette étude a donné lieu à une publication dans une conférence nationale [Meilender *et al.*, 2011b].

2.1 Les wikis

Les wikis sont des sites Web permettant la création et l'édition collaborative de contenus de manière simple [Wikipédia, 2013a]. La popularité de Wikipédia [Wikipédia, 2013b] montre l'importance de ce type de système. Après les avoir présentés (section 2.1.1), nous expliquons leur fonctionnement (section 2.1.2) et donnons quelques limites évoquées dans la littérature (section 2.1.3).

2.1.1 Présentation des wikis

L'idée de Ward Cunnigham lorsqu'il a introduit la notion de wiki en 1995 était de proposer le système de gestion de base de données le plus simple qui puisse fonctionner [Cunningham, 2002]. Ce principe fondateur constitue le principal vecteur de leur popularité. Les wikis peuvent être définis comme des sites Web permettant la création et l'édition collaborative de contenus en ligne [Leuf et Cunningham, 2001]. D'un point de vue pratique, ils sont généralement constitués d'un ensemble de pages éditables, organisées en catégories et reliées par des liens hypertextes. Ils sont devenus un des symboles de l'interactivité promue à travers le Web 2.0 : le contenu, la structuration, la mise en page et l'organisation des pages sont le fruit d'un travail collaboratif mené par une communauté d'utilisateurs.

De nombreux exemples de wikis ont vu le jour, le plus célèbre étant Wikipédia. En juillet 2012, sa version française comportait plus de 5 millions de pages élaborées par environ 1,3 millions de contributeurs. Cependant, limiter les wikis au phénomène Wikipédia serait restrictif. De nombreuses études ont montré l'apport des wikis dans différents domaines. Parmi ces travaux, des utilisations variées sont présentées, par exemple pour la gestion des ouvrages dans des bibliothèques médicales [Barsky et Giustini, 2007], en tant que support pédagogique pour l'éducation [Parker et Chao, 2007] ou dans la gestion de projet [Louridas, 2006].

2.1.2 Fonctionnement des wikis

D'un point de vue technique, les wikis sont créés et maintenus grâce à un type particulier de gestionnaire de contenu, le *moteur de wiki*. Depuis 1995, de nombreux projets de moteurs de wiki ont vu le jour. Le site WikiMatrix [CosmoCode, 2013] propose ainsi une comparaison des 136 moteurs de wiki qu'il recense. Chaque système dispose de ses propres particularités. Elles peuvent être fonction de l'usage auquel le moteur sera destiné : certains moteurs mettent l'accent sur les

```
"'D'Artagnan"', le héros populaire des "Trois Mousquetaires" d'[[Alexandre Dumas]], est finalement promu lieutenant des mousquetaires par le [[Cardinal de Richelieu]]. [[Catégorie:Personnage d'Alexandre Dumas]][[Catégorie:Naissance en Gascogne]]
```

(a) Extrait de wikitext d'une page sur d'Artagnan.

D'Artagnan, le héros populaire des *Trois Mousquetaires* d'**Alexandre Dumas**, est finalement promu lieutenant des mousquetaires par le **Cardinal de Richelieu**.
Catégories :**Personnage d'Alexandre Dumas**, **Naissance en Gascogne**

(b) Affichage final du texte sur la page du wiki.

FIGURE 2.1 – Extrait du wikitext (a) de *Mediawiki* d'une page sur d'Artagnan et son rendu final (b) .

relations sociales entre les membres (e.g. Socialtext, WikyBlog), tandis que d'autres proposent des structures généralistes (e.g. Mediawiki, le moteur de Wikipédia) ou spécialisées sur un domaine (e.g. Foswiki est spécialisé dans la gestion de projet). D'autres moteurs de wiki se distinguent par leur simplicité d'installation et d'utilisation (e.g. DokuWiki ne nécessite aucune base de données) ou des caractéristiques techniques telles que le langage de programmation (e.g. JSPWiki, écrit en Java grâce aux technologies JSP) ou l'intégration dans des solutions de gestion de contenu plus complètes (e.g. Drupal Wiki, prévu pour être intégré au gestionnaire de contenu Drupal).

Le travail collaboratif de création et de maintenance des contenus est rendu possible par l'emploi des *wikitexts*. Les wikitexts sont des langages de balisage choisis pour leur simplicité d'utilisation et leur faculté à inclure les fonctions nécessaires à l'édition des pages. Généralement, chaque moteur dispose de son propre wikitext, la figure 2.1 en est un exemple. Toutefois, certains sytèmes partagent les mêmes wikitexts ou permettent l'utilisation de wikitexts personnalisés.

2.1.3 Critiques et limites

Paradoxalement, le principal point fort des wikis est également le plus critiqué. En effet, l'ouverture des systèmes qui fait leur popularité constitue également une de leurs limites : en permettant l'édition de contenus par tous les utilisateurs, aucune garantie ne peut être donnée quant à la qualité du résultat final. Dans le cadre de Wikipédia, des travaux ont été menés pour limiter le vandalisme et un système de validation des contributions a été mis en place. De plus, les wikis sont le centre d'une communauté de contributeurs. Ainsi, les wikis spécialisés se construisent le plus souvent autour d'une communauté de spécialistes. Selon leur degré et leur domaine d'expertise, leurs interventions ne sont pas de la même qualité. Cette qualité varie également en fonction du niveau de maîtrise du système des contributeurs, d'autant plus qu'il n'existe pas d'outils permettant de l'estimer. Il conviendrait donc de travailler sur les différenciations au sein des communautés de contributeurs.

D'un point de vue fonctionnel, les systèmes de wiki reposent sur une architecture centralisée autour d'une base de données. Cela induit la mise en place de stratégies pour gérer les mises à jour concurrentes, c'est-à-dire des mises à jour simultanées dont le contenu est différent. Les wikis pair-à-pair répondent à cette problématique, à l'image du moteur Wooki [Weiss *et al.*, 2007].

Du point de vue de la gestion des connaissances, des limites ont déjà été détectées [Yu, 2011]. La première concerne la recherche d'informations. Dans de nombreux cas, des informations présentes dans le wiki ne peuvent être obtenues de manière automatique. Par exemple, aucune

requête sur Wikipédia ne peut donner la liste des personnages féminins des *Trois Mousquetaires* d'Alexandre Dumas, alors que cette information est présente sur la page wiki relative au roman. Le problème est que cette information est écrite en langue naturelle, donc uniquement accessible à un être humain lisant cette langue. Pour que cette information puisse être utilisée par une machine, un formatage spécifique est nécessaire. L'unique solution à ce jour pour les wikis est la multiplication des catégories et des listes. Mais il est inimaginable qu'une communauté crée automatiquement toutes les listes répondant à toutes les requêtes potentielles des utilisateurs.

La conséquence directe de la limite sur la recherche d'informations est l'impossibilité de les réutiliser. Ainsi de nombreuses informations sont « enfermées » dans un système et ne sont pas exploitables par les systèmes extérieurs. Cette situation va à l'encontre de la vision actuelle du Web de données, évoquée dans le chapitre 1.

Une autre limite concerne la cohérence des connaissances. L'ouverture du wiki permet de mélanger opinions et points de vue plus ou moins objectifs. Cela signifie qu'un sujet évoqué dans une page peut être évoqué de manière contraire dans une autre page du wiki sans que le système ne détecte d'incohérence. Cette situation peut être source de confusion pour les utilisateurs et nuire gravement à la crédibilité d'un système. La seule parade des wikis est la vérification et la comparaison des informations par les utilisateurs, mais les résultats sont d'autant plus difficiles à certifier que le wiki est grand.

2.2 Les wikis sémantiques

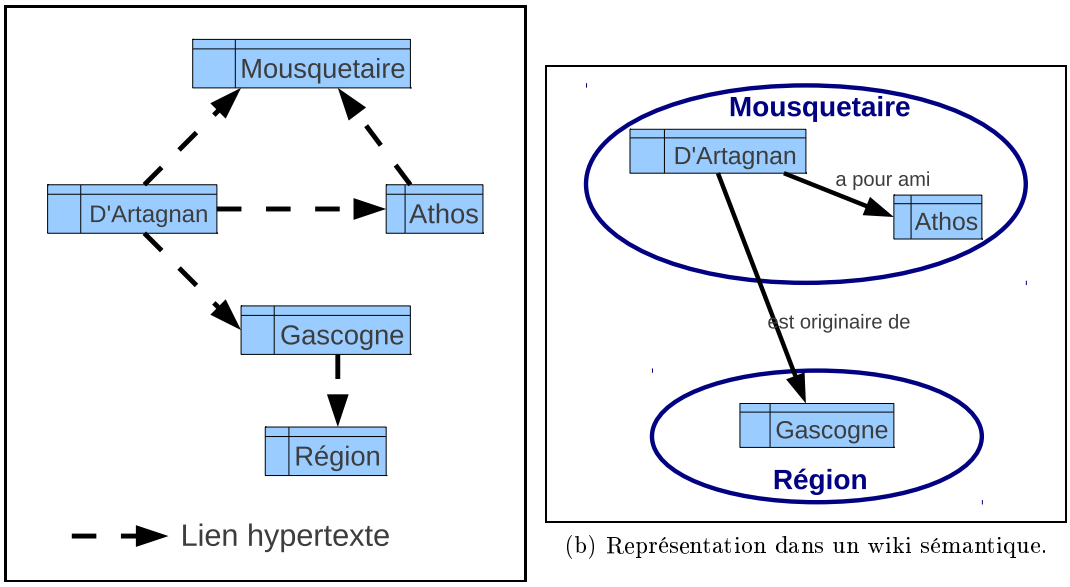
Plusieurs limites des wikis ont conduit à leur évolution, en particulier l'intégration des technologies du Web sémantique. Les wikis sémantiques sont définis dans la section 2.2.1. Ces wikis peuvent être caractérisés selon l'approche théorique qu'ils implémentent, soit *Wikis for Ontologies* soit *Ontologies for Wikis*. La section 2.2.2 présente ces approches tandis que la section 2.2.3 montre les apports des wikis sémantiques par rapport aux wikis traditionnels.

2.2.1 Qu'est-ce qu'un wiki sémantique ?

Les wikis sémantiques sont nés du rapprochement des wikis et du Web sémantique. Les wikis sémantiques sont définis par Berners-Lee et Fischetti [1999] comme des wikis améliorés par l'utilisation des technologies du Web sémantique. Plus particulièrement, un wiki sémantique est similaire à un wiki traditionnel dans le sens où c'est un site Web dans lequel le contenu est ajouté par les utilisateurs. Ce contenu est organisé en pages éditables et indexables, accessibles à tous les utilisateurs.

Cependant, contrairement au wiki traditionnel, le wiki sémantique ne se limite pas au texte en langage naturel. Il caractérise les ressources et les liens entre celles-ci comme l'illustre la figure 2.2. Ces informations sont formalisées et deviennent donc exploitables par une machine, à travers des processus de raisonnements artificiels. Ainsi, le but des wikis sémantiques est la création d'une ontologie ou le peuplement d'une ontologie préexistante, répondant aux standards du Web sémantique. Généralement, les wikis sémantiques considèrent chaque page comme une instance de l'ontologie et les catégories comme des classes. La sémantique des liens entre les pages est précisée, ce qui permet la création de propriétés.

Techniquement, les wikis sémantiques sont exploités grâce aux moteurs de wiki sémantique et utilisent généralement un wikitext adapté à la gestion des connaissances.



(a) Représentation dans un wiki classique.

(b) Représentation dans un wiki sémantique.

FIGURE 2.2 – Représentation d'une page concernant D'Artagnan dans un wiki classique (a) et dans un wiki sémantique (b).

2.2.2 *Wikis for Ontologies* et *Ontologies for Wikis*

Comme mentionné par Buffa *et al.* [2008], on peut historiquement distinguer deux approches dans la conception des moteurs de wikis sémantiques : *Wikis for Ontologies* (WfO) et *Ontologies for Wikis* (OfW). Si elle tend à disparaître, cette distinction est fondamentale pour comprendre la manière dont les moteurs de wikis gèrent l'ontologie sous-jacente, et comment cela se répercute sur le processus d'acquisition des connaissances. Les sections 2.2.2.1 et 2.2.2.2 présentent les deux approches, et la section 2.2.2.3 montre l'évolution des moteurs par rapport à ces notions.

2.2.2.1 *Wikis for Ontologies*

L'approche WfO (ou *Wikitology*) concerne le plus grand nombre des moteurs. Cette approche est centrée sur le texte : on cherche à conserver les textes présents dans les wikis classiques, en y ajoutant des annotations sémantiques. Dans ce cadre, les pages sont considérées comme des instances et les liens typés comme des propriétés. Il se dessine ainsi une ontologie formalisée, dont la précision est le plus souvent renforcée par la catégorisation des concepts. Dans ce cas, la structure des données est généralement très souple et permet une grande liberté pour l'utilisateur mais ne garantit pas l'utilisabilité des ontologies résultantes. En effet, elle suppose que l'utilisateur se rappelle au moment de la saisie de toutes les propriétés et de tous les concepts existants afin de ne pas créer de doublons. Par exemple, rien n'empêche un utilisateur de créer deux relations « a pour origine » et « est originaire de » ayant la même sémantique. Cela alourdit la base de connaissances et nuit à son homogénéité. D'un point de vue ergonomique, l'édition d'une page consistera à remplir une grande zone de texte libre. Le wikitext permet des annotations directement dans le texte, à la manière de ce qui est présenté à la figure 2.3.

2.2.2.2 *Ontologies for Wikis*

La deuxième approche, OfW, suppose généralement une ontologie pré-existante. Selon cette approche, l'objectif du moteur de wiki est de fournir les outils permettant son peuplement par l'ajout d'instances et parfois de classes. Ces outils sont généralement des formulaires à choix multiples ou utilisant l'auto-complétion. S'ils restreignent la liberté des utilisateurs, en refusant par exemple la création de nouveaux types de lien, ils garantissent la cohésion de l'ontologie finale. Cette approche apparente le wiki à un éditeur de métadonnées, permettant de peupler l'ontologie. Ainsi, les moteurs de wiki de cette catégorie sont le plus souvent destinés à des domaines spécifiques, plus facilement formalisables. Typiquement, lors de l'édition d'une page, l'utilisateur se voit proposer des formulaires où chaque question correspond à une propriété pré-existante dans l'ontologie.

2.2.2.3 Évolution des notions

Chacune des approches présentées précédemment propose des qualités et des défauts qui semblent s'opposer :

- l'approche WfO laisse une liberté totale à l'utilisateur, au détriment de la qualité de l'ontologie sous-jacente,
- l'approche OfW privilégie la cohérence de l'ontologie en guidant l'utilisateur, mais ne permet pas de conserver les informations non structurées ou non prévues explicitement dans l'ontologie éditée.

Dans les premières années des moteurs de wiki, cette distinction était facilement identifiable et permettait un classement strict des moteurs en deux catégories. La tendance a toutefois évolué ces dernières années. Il semble que les moteurs de chaque approche regardent dans le camp opposé pour s'en inspirer. Ainsi, par exemple, les moteurs actuels de l'approche WfO intègrent désormais systématiquement des formulaires pour contraindre et standardiser les annotations. Inversement, les moteurs de l'approche OfW intègrent de plus en plus souvent des espaces de texte libre. Celui-ci est généralement conservé sous forme de commentaires dans l'ontologie ou dans une base de données distincte.

2.2.3 Apports des wikis sémantiques

Pour pouvoir répondre aux attentes des utilisateurs, les moteurs de wiki sémantique doivent fournir des fonctionnalités avancées, en complément de celles proposées par les moteurs de wiki classique. Les sections suivantes montrent les améliorations attendues d'un moteur de wiki sémantique selon deux points de vue : celui de la gestion des connaissances et celui de l'ergonomie utilisateur.

2.2.3.1 Du point de vue de la gestion des connaissances

On attend d'un wiki sémantique qu'il soit capable de répondre à des requêtes complexes. Cela induit une recherche sémantique, et non uniquement une recherche de chaînes de caractères. La plupart des systèmes proposent ainsi la génération de listes de réponses, en s'appuyant sur un moteur d'inférences. Toutefois, dans de nombreux cas, la syntaxe des requêtes avancées est complexe pour un non expert.

De plus, ces wikis doivent s'intégrer dans des applications du Web sémantique. Pour cela, il faut que les développements proposent une compatibilité avec ses normes et ses standards. Ainsi,

la plupart des annotations se font par des triplets RDF, écrits selon des conventions différentes. La finalité est souvent de créer des ontologies RDF(S) ou OWL qui puissent être manipulées par d'autres applications. Ces ontologies peuvent également être consultées par des requêtes externes, de type SPARQL, si un point d'accès a été prévu. Les connaissances peuvent être stockées dans un système spécialisé, du type *RDFstore*.

Enfin, les systèmes doivent être garants de la qualité des informations qu'il contiennent. Ils peuvent alors fournir des outils pour vérifier la cohérence des connaissances, tels que des moteurs d'inférences.

2.2.3.2 D'un point de vue ergonomique

Un des principaux intérêts des wikis sémantiques est d'améliorer l'utilisation des wikis en exploitant les connaissances grâce aux technologies du Web sémantique. Ainsi, dans l'affichage des données, la navigation doit s'appuyer sur la sémantique. C'est le cas de certains systèmes proposant par exemple la navigation par facettes, qui guide l'utilisateur par la proposition de ressources pertinentes.

En termes d'utilisation, les wikis sémantiques deviennent des outils prisés pour l'annotation sémantique. En effet, leur édition peut être complexe. Toutefois, ce problème est avant tout ergonomique et de nombreux wikis proposent une aide à l'utilisateur. Alors que la plupart des systèmes s'appuient sur l'acquisition de triplets RDF, il peut paraître fastidieux pour un utilisateur non expert d'en comprendre les mécanismes. C'est pourquoi divers outils guident l'utilisateur dans l'édition, tels que des formulaires, des éditeurs WYSIWYG (*What You See Is What You Get*) ou de l'auto-complétion.

2.3 Les moteurs de wiki sémantique

Le site Semanticweb.org [2012] dénombre 39 projets de moteurs de wiki sémantique, dont une minorité est toujours en évolution. Cependant, plusieurs projets de moteur désormais inactifs ont présenté un apport majeur pour les wikis sémantiques. Ils sont évoqués dans la section 2.3.1. La section 2.3.2 présente les principaux moteurs actuels tandis que la section 2.3.3 fournit des éléments pour les comparer.

2.3.1 Retour sur quelques projets pionniers

Platypus wiki [Campanini *et al.*, 2004], qui implémente l'approche WfO, fut probablement le premier wiki sémantique. Les annotations sont séparées du texte et ajoutées par un éditeur distinct de celui de la page wiki. Elles peuvent être faites soit en OWL, soit en RDF(S) et sont considérées comme des metadonnées. Platypus propose des outils pour éditer l'ontologie, mais ne propose pas de moteur d'inférences. La cohérence des données n'est donc pas garantie. Il ne propose pas non plus de système de requête sur les annotations, qui sont uniquement utilisées pour améliorer la navigation.

Rhizome [Souzis, 2006] propose une approche technique originale. Il utilise certaines applications de XSLT en RDF comme RxSLT pour le formatage du texte et RxPath pour le langage de requête. D'un point de vue conceptuel, il s'appuie sur ZML, un wikitext spécifique proche de RDF permettant d'identifier simplement les propriétés dans la page. Cependant, il ne dispose que de peu de fonctionnalités. En effet, les pages sont stockées en RDF et peuvent être modifiées

par un autre système, mais il n'y a pas de système de contrôle de la cohérence de l'ontologie. En outre, il ne dispose pas de système de requêtes avancées, c'est-à-dire de requêtes exploitant des raisonnements. Néanmoins, Rhizome propose une gestion complexe des droits des utilisateurs, avec des mécanismes d'autorisation et de validation.

WikSAR [Aumüller et Auer, 2005] (dans la continuité de **SHAWN**) propose une avancée intéressante dans le mode d'édition. En effet, il permet l'édition de triplets RDF avec un langage wikitext facilement compréhensible par un utilisateur. Il en extrait une ontologie pour laquelle il fournit un outil spécifique de navigation mais dont il ne garantit toutefois pas la cohérence. Il utilise les metadonnées pour proposer une navigation par facettes et permet des requêtes avancées du type « Qui est l'auteur de *Hamlet* ? ».

Kaukulo [Kiesel, 2006] implémente l'approche *Wikitology*. Il permet l'édition de triplets RDF, en proposant un intéressant mécanisne d'auto-complétion. On peut noter qu'il est le premier (et un des seuls) à proposer l'import d'ontologies RDF(S).

COW [Fischer *et al.*, 2006] utilise une approche similaire. Il propose un stockage séparé de l'ontologie avec une gestion de ses versions. Il introduit également le concept de modèle de requête qui permet de lister automatiquement des éléments dans une page grâce à une requête insérée lors de l'édition.

De nombreux wikis implémentent l'approche *ontologies for wikis*. Parmi ceux-ci, notons **Sweetwiki** [Buffa *et al.*, 2008]. Ce wiki permet de créer et d'annoter collectivement des pages qui peuvent contenir différents types de ressources comme du texte, des photos ou des vidéos. Il se différencie des wikis précédents sur de nombreux points. Tout d'abord, les développeurs ont eu la volonté d'utiliser des standards du Web (XHTML, XML) et du Web sémantique (RDFa, RDF(S), OWL, SPARQL). Ensuite, Sweetwiki cherche à identifier des communautés d'utilisateurs et, en ce sens, construit une passerelle vers le Web social. Enfin, il propose de nombreux outils d'édition tels qu'un éditeur WYSIWYG, un système d'auto-complétion et un éditeur léger d'ontologies.

2.3.2 Projets actifs de moteurs

2.3.2.1 Semantic MediaWiki

Principe de fonctionnement. Probablement le plus populaire et le plus utilisé des moteurs de wiki sémantique, Semantic MediaWiki [Semantic MediaWiki, 2013] (SMW) s'appuie sur MediaWiki [MediaWiki, 2013] (MW), le moteur de Wikipédia. La création du moteur est d'ailleurs intimement liée à l'encyclopédie collaborative. En effet, la première vision du projet propose une amélioration de Wikipédia prenant en compte les technologies du Web sémantique en ajoutant la notion de liens typés [Krötzsch *et al.*, 2005]. L'idée est approfondie par Krötzsch *et al.* [2007b] en intégrant les notions de catégories, d'attributs et de types. De cette vision est née SMW. Le but de cette extension est d'ajouter une couche sémantique à MW, qui se place en complément des fonctionnalités existantes. Le moteur permet donc d'ajouter des annotations à chaque page, sans modifier les parties non structurées déjà présentes [Krötzsch et Vrandecic, 2011].

SMW+ [Semanticweb.org, 2013] est une version étendue de SMW, intégrant des extensions disponibles en *Open Source*. Développé initialement par la défunte société allemande Ontoprise, le projet semble avoir été repris par la société DIQA sous le nom de DataWiki [DIQA, 2013].

(a) Exemple d'une page de wiki sur d'Artagnan.

```
"'D'Artagnan"' vient de [[Gascogne]]. [[Catégorie:Mousquetaire]]
```

(b) Wikitext de la page précédente dans MW.

```
"'D'Artagnan"' vient de [[est originaire de::Gascogne]].
[[Catégorie:Mousquetaire]]
```

(c) Wikitext de la page précédente dans SMW.

FIGURE 2.3 – Wikitext d'une page concernant d'Artagnan (a) dans MW (b) et dans SMW (c).

Édition des connaissances. Afin de préserver la simplicité et les habitudes des utilisateurs, le langage d'annotation de SMW étend le wikitext de MW. La figure 2.3 montre un exemple de comparaison des wikitexts dans MW et dans SMW, ainsi que la manière dont le moteur peut les interpréter. Chaque entité de l'ontologie correspond à une page dans le wiki et des espaces de nom sont utilisés pour les différencier. Dans chaque page, les annotations ne peuvent se rapporter qu'à cette entité. Ainsi la majorité des pages correspond à des instances. Ces instances peuvent être classifiées selon des concepts correspondant aux catégories dans MW, et peuvent être reliées par des propriétés dont on peut définir le co-domaine dans les pages correspondantes. Ainsi, le modèle de SMW est très proche de OWL bien qu'il n'exploite qu'un fragment de l'expressivité de ce langage. Ce mode d'édition confère à SMW une approche très ouverte de l'édition des connaissances : ainsi, il est possible d'utiliser des concepts ou d'ajouter des annotations exploitant des ressources qui ne sont pas encore définies dans l'ontologie au moment de l'édition d'une page.

Afin de simplifier l'édition, SMW a adapté l'utilisation des modèles de MW qui permettent de créer des structures applicables à plusieurs pages. Il est également possible d'utiliser des extensions telles que des formulaires ou des éditeurs WYSIWYG. Smart *et al.* [2009] proposent ainsi une interface pour la saisie des annotations dans un langage naturel contrôlé, proche de ce qui est proposé par AceWiki, le moteur présenté à la section 2.3.2.4.

Navigation et interrogation des données. La présence de liens typés entre les pages permet la navigation dans l'ontologie sous-jacente, en particulier en affichant des *factbox* dont l'objet est de recenser toutes les annotations sémantiques d'une page. La navigation peut être améliorée dans SMW par l'utilisation des nombreuses extensions répertoriées sur le site institutionnel, comme par exemple la navigation par facettes.

L'interrogation des données peut se faire à travers l'utilisation d'un langage de requête en

ligne. Ce langage est très simple d'utilisation mais présente une expressivité limitée, équivalente à la logique de descriptions $\mathcal{EL}^{++}$. Des extensions de SMW permettent d'intégrer un *triple store* et un point d'accès SPARQL. Il est intéressant de noter que le moteur d'inférences KAON2 [KAON2, 2013] est développé en collaboration avec SMW et s'intègre donc facilement dans le *framework*.

Exemples d'applications. De nombreuses applications de SMW ont été décrites dans la littérature. Une expérience sur la création d'un portail sémantique sur un institut de recherche académique est rapportée par Herzig et Ell [2010]. Les annotations sémantiques sont utilisées pour décrire les composantes de l'institut et le personnel, et le wiki sémantique permet des alignements avec des ontologies externes telles que *FOAF*. L'article souligne l'importance des modèles et propose une réflexion sur l'introduction de données structurées dans les pages de texte libre. Des exemples de requêtes en ligne sont également présentés. Le wiki sémantique final contient environ 16 700 pages décrites par plus de 105 000 annotations.

Dans un autre contexte, *Wikitaaable* [Badra *et al.*, 2008a; Cordier *et al.*, 2009] est un wiki sémantique utilisé dans la gestion de recettes de cuisine. La finalité de ce wiki est principalement d'enrichir la base de connaissances grâce à une communauté d'utilisateurs. Ceux-ci peuvent consulter, modifier ou ajouter des recettes de cuisine formalisées, en les reliant avec une hiérarchie d'ingrédients. Ces recettes peuvent être annotées et indexées, et rendues utilisables par un moteur de raisonnement à partir de cas qui peut proposer des adaptations.

2.3.2.2 KiWi

Principe de fonctionnement. KiWi[33] [KiWi, 2012] peut être vu comme un compromis entre les approches WfO et OfW. Ce moteur propose d'intégrer d'une part l'édition de texte non structuré à la manière d'un wiki classique et d'autre part l'édition d'annotations sémantiques. Le système propose deux types d'édition distincts, par wikitext pour le texte non structuré et par formulaire pour les annotations sémantiques. Ainsi, KiWi a été conçu comme un outil collaboratif à destination d'utilisateurs de différents niveaux en ingénierie des connaissances. KiWi repose sur une ontologie préexistante, chargée au moment de l'installation, qui est peuplée lors de l'édition des annotations sémantiques [Schaffert, 2006]. Les différents niveaux de structuration sont stockés séparément et synchronisés grâce à la combinaison d'une base de données relationnelle et d'un *triple store* [Sint *et al.*, 2009].

Édition des connaissances. L'édition dans KiWi se passe de manière séparée pour le texte de la page et les annotations syntaxiques. Le texte est édité de manière classique dans un éditeur WYSIWYG. Dans le cas des annotations sémantiques, la saisie se fait à l'aide de formulaires. Chaque page dans KiWi correspond à une ressource de la base de connaissances et est associée à au moins un concept. Ainsi, par défaut, KiWi propose à l'utilisateur des annotations correspondant au concept. La figure 2.4 montre des formulaires d'édition d'annotations syntaxiques sur une page. Un éditeur d'annotation permet d'ajouter des concepts à une page. Cette approche, bien plus stricte que celle de SMW, permet de faciliter la maintenance de la base de connaissances et garantit l'absence d'incohérence dans l'ontologie. Toutefois, le système de *versioning* présent dans la plupart des moteurs ne s'applique qu'à la partie texte, et non aux annotations. En plus des mécanismes d'édition, KiWi fournit des extensions qui permettent de modifier l'ontologie préexistante, par exemple en ajoutant des sous-concepts et des super-concepts, ou en précisant les domaines et co-domaines des propriétés.

33. *Knowledge in Wiki*, précédemment *IkeWiki*.

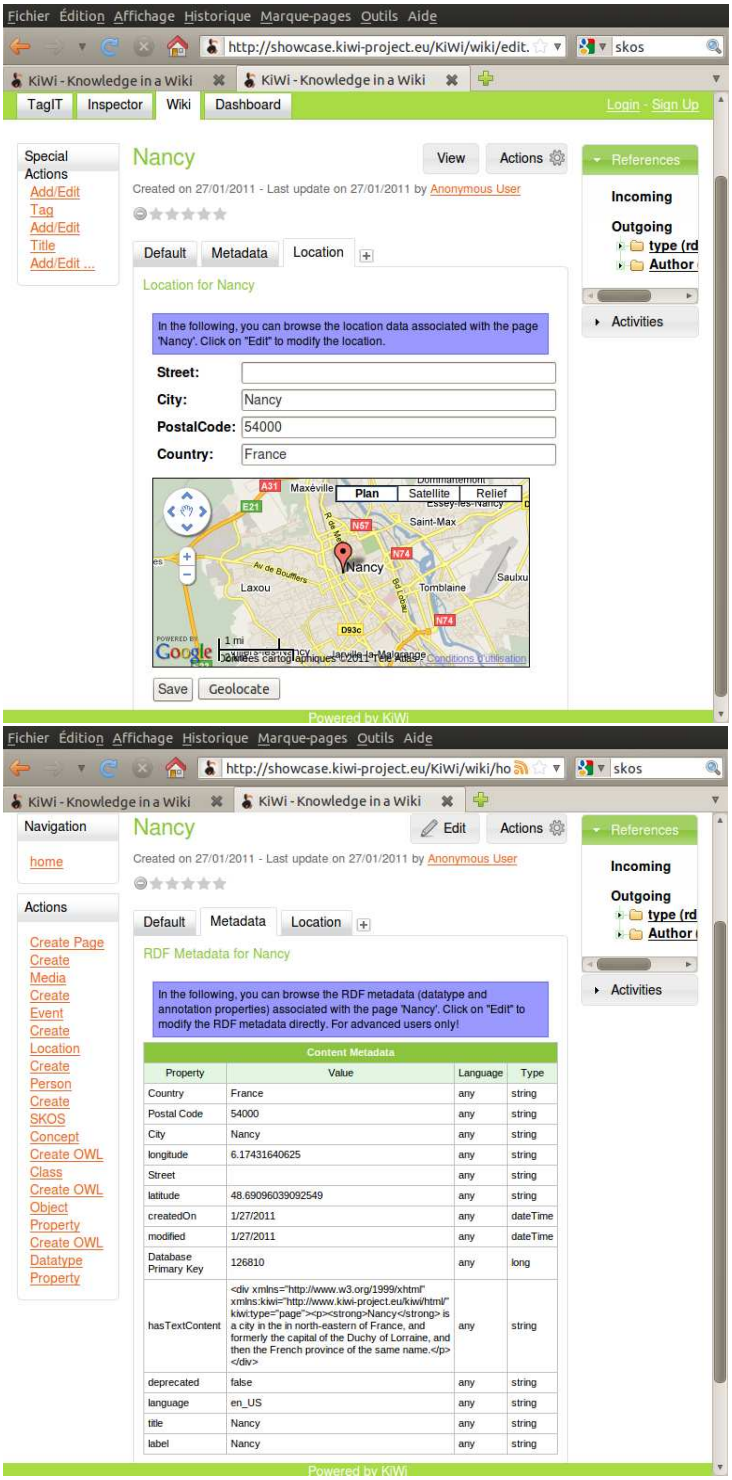


FIGURE 2.4 – Formulaires d'édition d'une instance sous KiWi.

Navigation et interrogation des données. L'interface de KiWi repose sur des technologies AJAX. La sémantique est exploitée pour améliorer la navigation, par exemple en associant un concept à une vue spécifique. Une grande attention a été portée à la gestion des fichiers multimédia.

De manière interne et externe, KiWi permet l'emploi de requêtes SPARQL, ce qui permet d'obtenir des requêtes internes plus expressives que dans SMW [Krötzsch *et al.*, 2007a]. Les raisonnements sont effectués par le moteur d'inférences Pellet [Sirin *et al.*, 2007b].

Exemples d'applications. Quelques exemples de projets d'application sont donnés sur le site institutionnel du moteur, à travers la description de scénarios d'utilisation. Ces applications sont liées directement aux partenaires institutionnels du projet KiWi mais, à notre connaissance, aucune information n'est donnée dans la littérature quant à leur réalisation.

2.3.2.3 OntoWiki

Principe de fonctionnement. OntoWiki [OntoWiki, 2012] est l'exemple typique de moteur de wiki sémantique implémentant l'approche OfW. Le principal objectif d'OntoWiki est de simplifier la représentation et l'acquisition d'instances. Contrairement aux moteurs de l'approche WfO, le moteur ne permet pas de mélanger l'édition de texte et les annotations sémantiques [Auer *et al.*, 2006]. Sa philosophie est d'appliquer à l'édition de connaissances le paradigme des wikis de « faciliter les corrections d'erreurs, plutôt que de les rendre difficiles à commettre »[34]. Les utilisateurs peuvent ainsi compléter des descriptions d'instances en respectant le schéma de l'ontologie. Ainsi, OntoWiki permet d'élaborer collaborativement des ontologies RDF(S) et OWL. Le moteur de wiki peut importer une ontologie existante pour la compléter ou permettre de créer une ontologie *ex nihilo*.

Représentation des connaissances. La représentation des triplets RDF de l'ontologie se fait selon des « cartes d'information »[35]. Chaque ressource RDF dispose d'une telle carte qui est représentée dans une page HTML, contenant des liens hypertextes vers les ressources utiles à la description. De plus, pour certaines instances, la page HTML peut être adaptée. Ainsi, si une instance contient des valeurs de propriétés représentant des informations géographiques, une carte géographique est dynamiquement proposée. De même, une vue de type calendrier est proposée si certaines valeurs de propriété ont le type de données *xsd:date*.

Édition. OntoWiki propose de nombreux formulaires pour l'édition des connaissances. Chaque élément du formulaire représente une propriété de l'ontologie applicable à l'instance éditée. L'utilisateur est guidé pour le type de valeur à renseigner, le système se fondant sur le co-domaine de la propriété à renseigner. Un exemple de formulaire d'édition est montré à la figure 2.5.

Navigation et interrogation des données. OntoWiki propose de nombreux outils pour faciliter la navigation dans l'ontologie comme des vues permettant de visualiser la hiérarchie des concepts, un système de navigation par facettes, la recherche textuelle et des listes utiles telles que la liste des instances d'un concept ou la liste des propriétés applicables à une instance. Ces systèmes de navigation peuvent être combinés pour affiner les résultats. Ils reposent tous sur

34. « *making it easy to correct mistakes, rather than making it hard to make them* » [Leuf et Cunningham, 2001].
35. « *information maps.* »

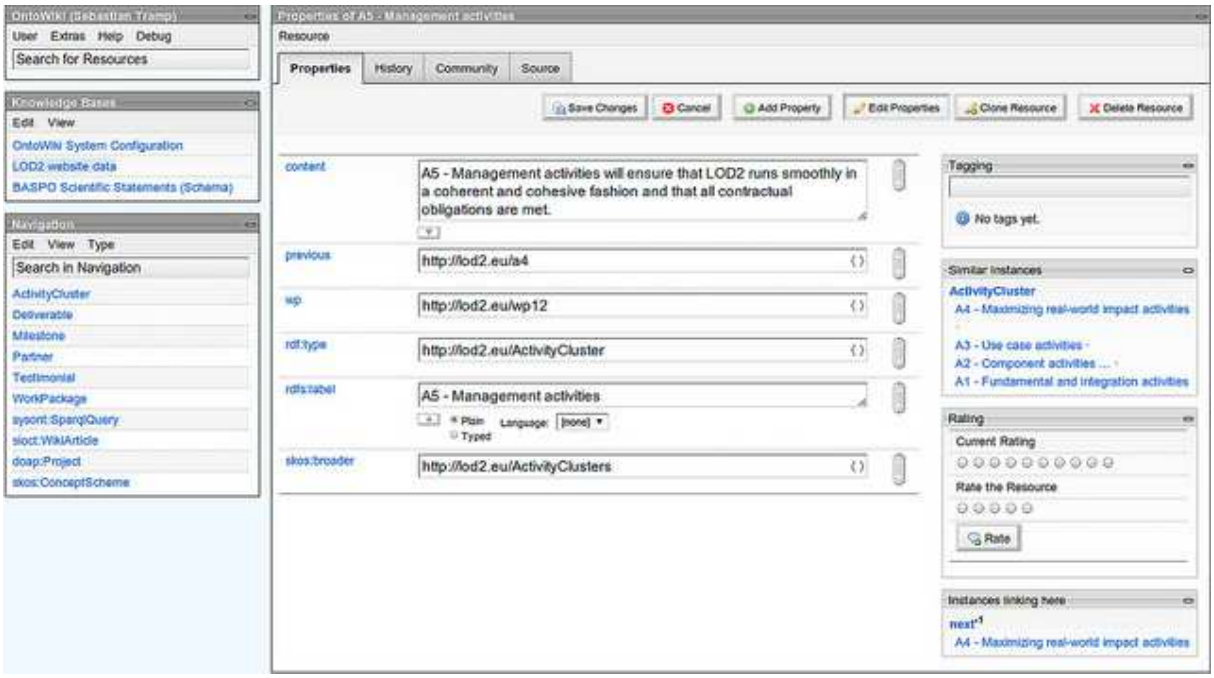


FIGURE 2.5 – Formulaire d'édition d'une instance sous OntoWiki.

des requêtes SPARQL. OntoWiki propose également un support original des contenus multilingues [Martin *et al.*, 2011].

Le moteur est spécialement pensé pour être connecté au Web de données. Il dispose d'un point d'accès SPARQL et permet l'import d'autres sources de données, pouvant agir aussi bien comme un serveur que comme une source de données liées [Tramp *et al.*, 2010].

Applications. Dans la littérature, quelques exemples d'applications sont donnés. OntoWiki sert de base à une étude prosopographique se fondant sur le *Catalogus Professorum Model*, un ensemble de vocabulaires et d'ontologies concernant les professeurs de l'Université de Leipzig [Riechert *et al.*, 2010]. Le contenu du wiki a été édité directement par des historiens, soutenus par des ingénieurs des connaissances et des experts du Web sémantique. Le résultat est consultable en ligne sous forme de Linked Data [Universität Leipzig, 2013]. D'autres applications sont évoquées, telles que la gestion d'une base de connaissances d'environ 240 000 triplets sur les araignées caucasiennes [Heino *et al.*, 2009] ou une base de connaissances des informations touristiques néerlandaises [Martin *et al.*, 2011].

2.3.2.4 AceWiki

Fonctionnement et représentation des connaissances. Le but d'AceWiki [Kuhn, 2009] est d'implémenter l'approche OfW en proposant des outils simples et novateurs pour le peuplement de l'ontologie. Pour cela, AceWiki propose une syntaxe spéciale reposant sur le langage ACE (*Attempto Controlled English*) [de Coi *et al.*, 2009]. La figure 2.6 donne un exemple de page éditée dans ce langage. ACE est un langage naturel contrôlé, au sens où c'est un fragment de l'anglais qui peut être directement formalisé. Le principe d'ACE est de représenter des prédicats formels d'une manière qui ressemble au langage naturel de façon à les rendre plus compréhensibles

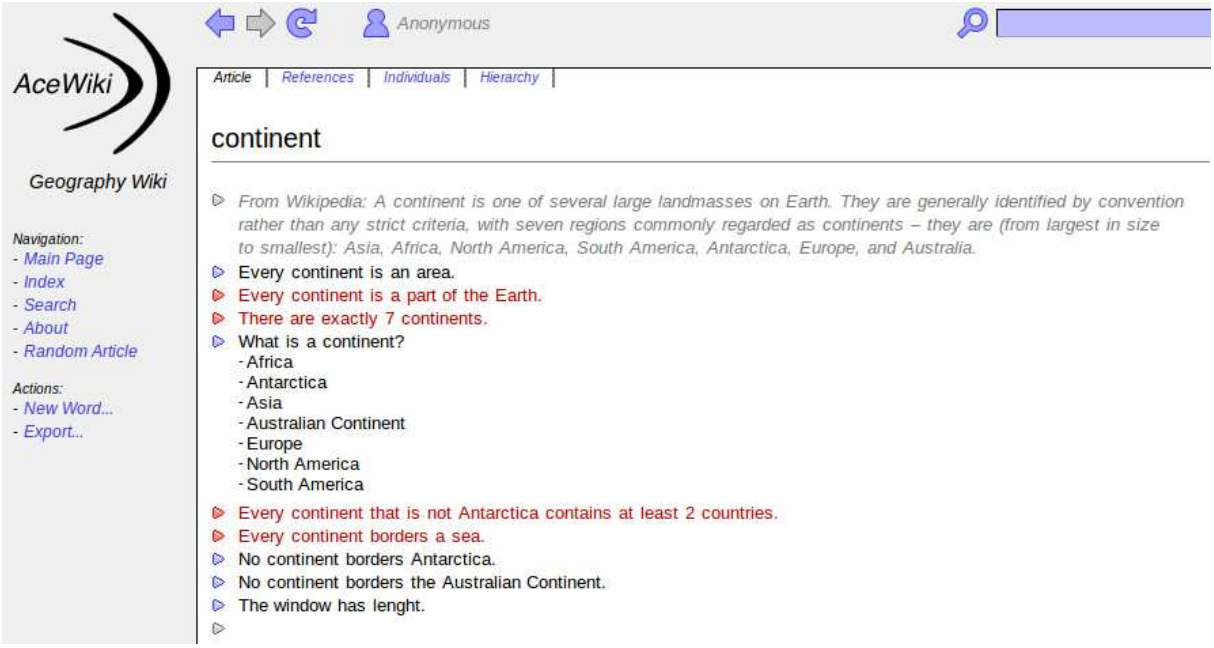


FIGURE 2.6 – Une page sur les continents éditée avec le langage ACE dans AceWiki.

aux personnes n'ayant aucune expertise en ingénierie des connaissances. Ainsi chaque prédicat peut être traduit dans des langages logiques, en particulier dans les langages du Web sémantique OWL et SWRL. Par exemple, le prédicat ACE

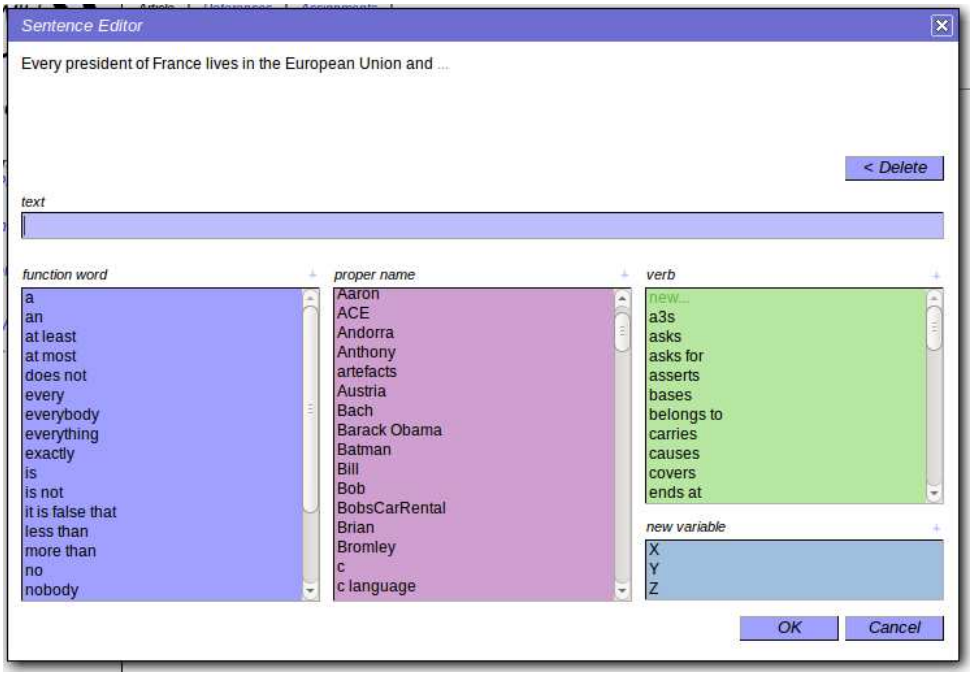
Every employee that does not own a car owns a bike.

peut se traduire dans la logique de descriptions $\mathcal{SROIQ}(D)$de la manière suivant :

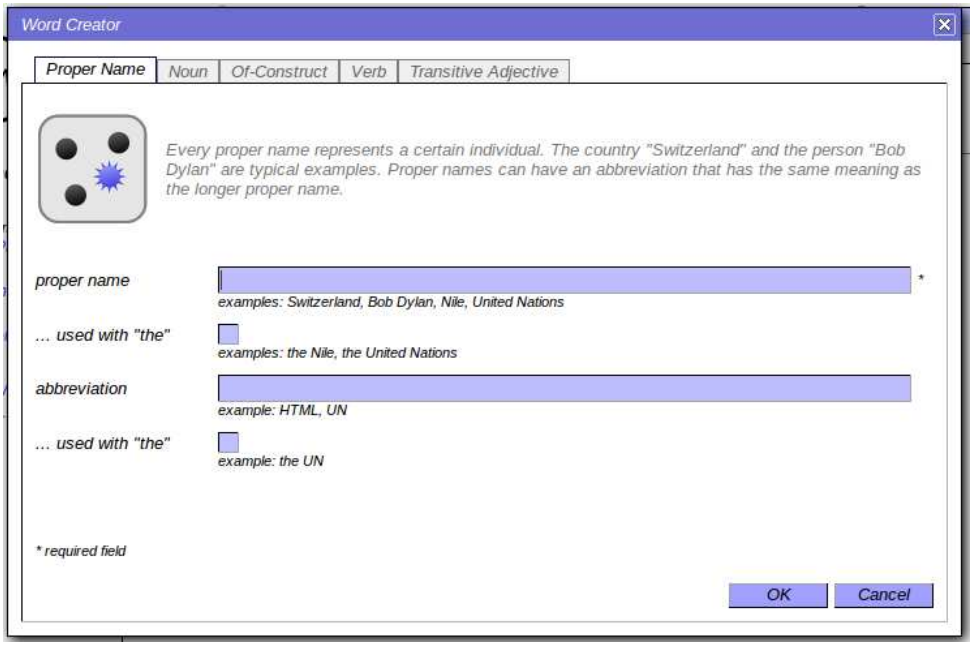
$$\texttt{Employee} \sqcap \neg(\exists \texttt{own.Car}) \sqsubseteq \exists \texttt{own.Bike}$$

Des tables de correspondance permettent de créer une correspondance entre les structures de phrases ACE et les axiomes, les concepts et les propriétés OWL. Par exemple, dans une phrase ACE, les termes *thing* et *something* correspondent au concept *owl:Thing*. De même, les noms propres issus du vocabulaire ACE, comme par exemple des noms de pays, sont interprétés comme des individus de l'ontologie, les noms communs comme des concepts, etc. Toutefois, un bon nombre de phrases d'ACE ne peuvent pas être traduites en OWL car le langage contrôlé permet une plus grande expressivité qu'OWL DL d'après les concepteurs. Dans ce cas, les phrases ne sont pas considérées dans l'ontologie.

Édition des connaissances. La simplicité du wiki repose sur le langage ACE, proche du langage naturel. Il est donc nécessaire de guider l'utilisateur dans l'apprentissage de ce langage, qui, s'il est compréhensible, ne correspond pas tout à fait au langage naturel. Ainsi, AceWiki propose un éditeur prédictif pour aider à la création et à la modification de phrases ACE, comme le montre la figure 2.7a. Il assure la cohérence lexicale et grammaticale des phrases en proposant à l'utilisateur les mots du vocabulaire qui semblent correspondre le mieux au contexte de la phrase. Ces mots sont proposés en s'appuyant sur les hiérarchies de rôles et de concepts de l'ontologie, ainsi que sur les domaines et les co-domaines des propriétés. De plus, l'éditeur d'AceWiki propose un ensemble de modèles de phrases à l'utilisateur afin de le guider dans les



(a) Ajout de phrases en langage ACE dans une page.



(b) Création d'une page dans AceWiki.

FIGURE 2.7 – Éditeurs d'AceWiki pour l'ajout d'une phrase en langage ACE (a) et la création d'une page (b).

constructions syntaxiques. Enfin, lors de la création des pages, l'utilisateur est invité à préciser le type de l'élément à ajouter (nom propre, nom commun, etc.) afin de respecter la syntaxe d'ACE. Un aperçu du formulaire utilisé est montré dans la figure 2.7b.

Raisonnement. AceWiki utilise le moteur d'inférences HermiT [Glimm *et al.*, 2010] pour raisonner dans le wiki avec les phrases qui peuvent être traduites en OWL. Afin de pouvoir assurer la cohérence de l'ontologie, chaque phrase est testée immédiatement après sa saisie dans le wiki. Si elle comporte une incohérence ou en introduit une dans la base de connaissances, alors elle n'est pas introduite dans l'ontologie et est affichée dans la page avec la couleur rouge. Le moteur d'inférences est également utilisé automatiquement pour inférer l'appartenance des instances aux concepts existants et pour construire la hiérarchie des concepts. Enfin, le moteur peut également être utilisé pour répondre à des requêtes présentes dans les pages en langage ACE.

Applications. Jusqu'à présent, AceWiki souffre d'un manque d'applications réelles. Seules trois sont laissées en lignes afin de servir de démonstration pour le système, accessibles depuis la page d'accueil du site Internet du système [AceWiki, 2013]. L'une concerne le langage ACE lui-même, une autre propose une application d'AceWiki dans le domaine de la géographie, tandis que la dernière est un mode « bac à sable » ouvert aux utilisateurs.

2.3.2.5 KnowWE

Principe de fonctionnement. KnowWE est une solution commerciale et *open source*. La vision première des développeurs de KnowWE était de construire un moteur extensible qui puisse s'adapter au domaine qu'il souhaite modéliser [Reutelshöfer *et al.*, 2009]. L'idée est donc de proposer un *framework* dans lequel puissent s'intégrer différents modules complémentaires à la gestion des connaissances, proposant par exemple des outils de traitement automatique du langage ou de fouille de textes.

Édition des connaissances. Les dernier travaux de développement ont mis en application ce principe, en spécialisant KnowWE à la représentation des connaissances liées à la résolution de problèmes [Baumeister *et al.*, 2011b]. Ainsi, chaque page représente un problème : la première partie de la page contient du texte pouvant être annoté à la manière de Semantic Mediawiki (avec toutefois un wikitext plus étendu) et la seconde partie est consacrée à l'édition d'une base de connaissances pour la résolution de ce problème. Cette seconde partie peut être éditée de deux manières : soit par l'emploi d'un langage de règles *ad hoc* propre au système, soit par l'édition d'arbres de décision. Chaque nœud de l'arbre correspond à un terme du vocabulaire préalablement édité dans une ontologie sous-jacente. La base est ensuite compilée au format propriétaire *d3web* [d3web, 2013] de manière à pouvoir être utilisée par un système d'aide à la décision dédié. La figure 2.8 présente une page comportant ces deux niveaux d'édition.

Raisonnement. Les deux niveaux d'édition impliquent deux niveaux de raisonnement. Le premier niveau concerne les annotations en wikitext : dans ce cas, les connaissances sont interrogeables de manière externe et interne par l'emploi de requêtes SPARQL. Dans le second niveau, l'emploi de raisonnement se fait par l'intermédiaire d'un moteur propre à d3web vérifiant la cohérence de celle-ci. La base peut être interrogée sur le wiki par l'intermédiaire d'un formulaire.

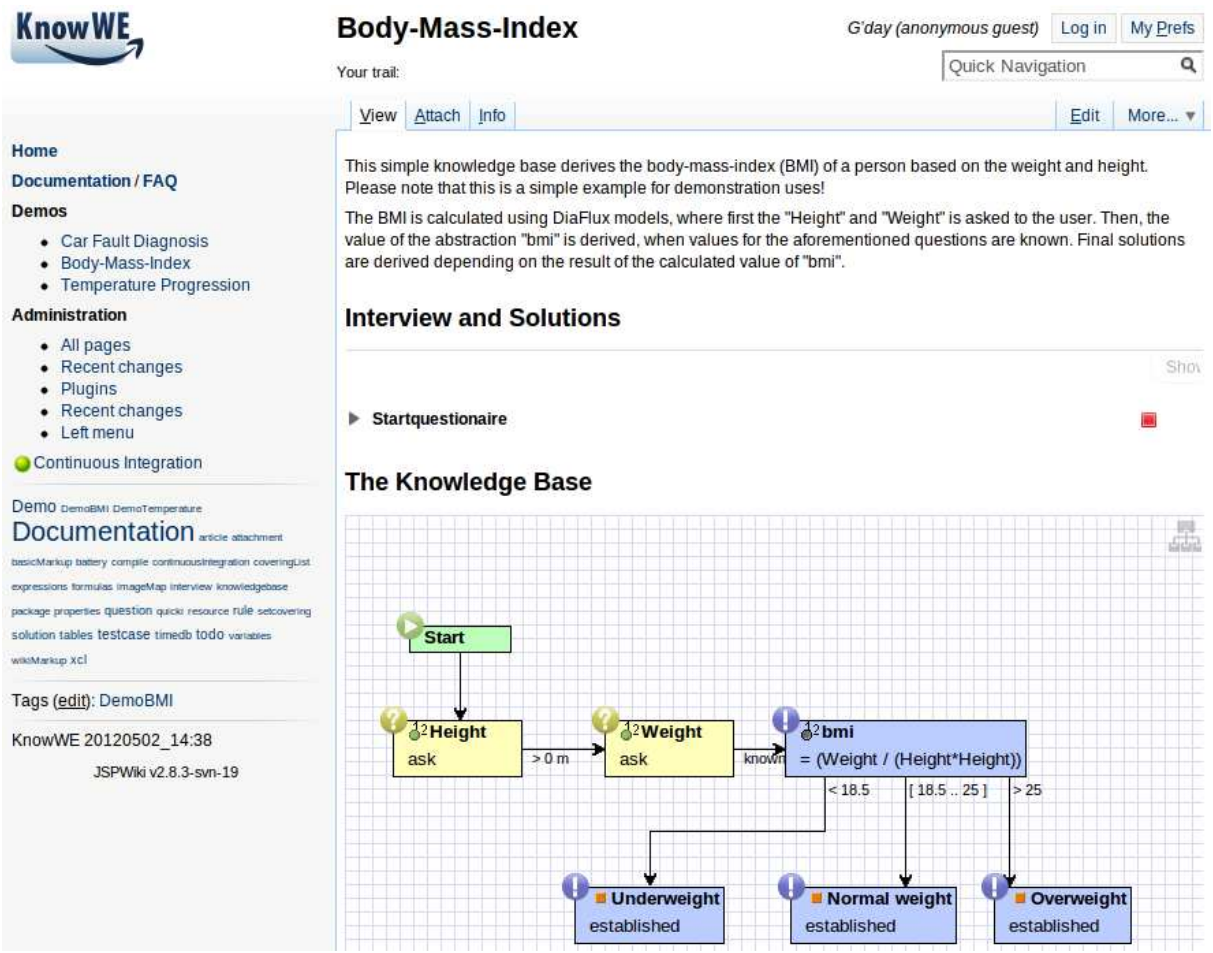


FIGURE 2.8 – Page concernant l'indice de masse corporelle sur KnowWE.

Applications. Plusieurs applications sont présentées dans la littérature. Hatko *et al.* [2010] évoquent l'édition d'un GBPC sur le sepsis[36]. Il est intéressant de noter que cette expérience ne concerne qu'un petit groupe d'experts rassemblé dans une même pièce, communiquant les éditions à un ingénieur des connaissances. Cela est justifié par les concepteurs du système par le fait que ce moteur de wiki n'est pas destiné aux grandes communautés. Une plus grande application commerciale dans le domaine médical est présentée par Baumeister *et al.* [2011b]. Le but du projet est de proposer des outils de questions/réponses pour les équipes d'urgence mobiles. Dans ce cas, des référentiels ont été édités suivant les consignes officielles des autorités allemandes.

2.3.2.6 Autres systèmes

Des wikis sémantiques spécialisés dans certains domaines ont également été créés comme **BOWiki** [Bacher *et al.*, 2008] et **SWiM** [Lange, 2008]. L'objectif de BOWiki est de proposer un outil pour la construction collaborative d'ontologies biomédicales. Lors de l'ajout d'une unité de connaissances, il contrôle la cohérence de l'ontologie grâce à un moteur d'inférences et la

36. Un sepsis est une infection générale grave de l'organisme par des germes pathogènes.

rejette le cas échéant. Le moteur d'inférences est également utilisé pour répondre à des requêtes complexes. L'idée est de contrôler la connaissance avant de la placer dans l'ontologie, c'est-à-dire de compléter le travail de l'expert par une vérification de la cohérence. En ce qui concerne SWiM, il implémente IkeWiki en ajoutant un support du langage OpenMath. Ce dernier est un langage XML de représentation des équations. Le but est d'en extraire la sémantique afin de faire de SWiM un outil capable de gérer l'édition collaborative de définitions de symboles mathématiques et leur apparence.

Certains moteurs proposent un mode de représentation des connaissances différent. Assez peu référencée et documentée, **Subleme** [Subleme, 2013] est une application minimaliste pour le partage d'informations. Tout y est considéré comme des pages (documents texte, utilisateurs, niveaux d'accès, etc.) reliées par des liens typés. Il n'utilise pas les notions de concepts ou d'instances mais gère des relations du type « est instance de ». Bien plus complet, **TaOPis** [Schatten *et al.*, 2009] propose l'édition d'ontologies sur un modèle orienté objet. Pour ce faire, il utilise la logique décidable F-logic. Il intègre un moteur d'inférences spécifique, Flora-2, qui utilise une syntaxe qui lui est propre pour les requêtes avancées.

Quelques systèmes commerciaux sont apparus dans la famille des wikis sémantiques dont **Knoodl** [Knoodl, 2013] et **Wikidsmart** [zAgile, 2013] sont des exemples. Knoodl se présente comme un outil de développement communautaire d'ontologies OWL et de bases de connaissances. Il dispose de fonctionnalités d'import et d'export d'ontologies, propose la gestion de requêtes avancées et un point d'accès SPARQL. L'idée de communautés est mise en avant par un système dans lequel chacune d'entre elles dispose de son propre wiki, intégrant son vocabulaire spécifique. Pour sa part, Wikidsmart entre dans le cadre d'une application Web sémantique plus lourde dont le but est de proposer des services pour la gestion commerciale et humaine des entreprises. Il propose le même type de solution que son concurrent. Dans les deux cas, il semble que les fonctions les plus intéressantes soient payantes, et les sociétés restent discrètes sur les moyens mis en œuvre.

2.3.3 Comparaison des systèmes

Pour comparer les différents projets actifs, nous nous sommes appuyés sur plusieurs critères, répartis dans quatre catégories.

La première catégorie concerne l'approche conceptuelle où nous cherchons à savoir si les concepteurs se fondent sur une approche *Wikis for Ontologies* ou une approche *Ontologies for Wikis*.

La deuxième catégorie permet de juger de l'emploi des technologies du Web sémantique, en s'appuyant en premier lieu sur le langage de représentation des connaissances utilisé. Nous cherchons à savoir ensuite dans quelle mesure l'import et l'export d'ontologies sont supportés. Enfin, nous nous intéressons aux méthodes de raisonnement et aux possibilités pour les requêtes internes (type de requêtes acceptées) et externes (disponibilité d'un point d'accès SPARQL).

La troisième catégorie traite de l'utilisabilité du système, en s'appuyant sur des critères propres aux wikis. Le mode d'annotation, les modes d'édition (éditeur WYSIWYG, formulaires, etc.), la gestion des droits des utilisateurs et la gestion des versions y sont recensés.

Enfin, la dernière partie regroupe des informations complémentaires sur les projets telles que le langage de programmation utilisé, la licence et la stabilité du système.

Le tableau 2.1 confronte les systèmes avec les critères précédemment énoncés. Il est intéressant de noter que, d'un point de vue global, SMW+, KiWi et Knoodl sont les plus complets. Les deux premiers sont les références des approches conceptuelles qu'ils implémentent. On remarque

	Approche		Outils du Web sémantique								Fonctions wiki							Statut		
	Wiki for ontology	Ontology for wiki	Langage de représentation	Export ontologie	Import ontologie	Système de requête	SparQL endpoint	Moteur inférence	Stockage des données	Annontations	Syntaxe	WYSIWYG	Formulaire	Autre mode d'édition	Gestion version	Gestion utilisateurs	Langage	Licence	Version
AceWiki		•	OWL, SWRL	○	○	Texte, requêtes avancées	○	HermiT	Texte	Dans le texte	ACE	○	○	Éditeur prédictif	○	○	Java	LGPL	alpha
BoWiki		•	OWL	●	OBO format	Texte, requêtes avancées	○	Pellet	DB	Inconnu	wikitext	○	○	○	●	●	PHP	GPL	stable
KiWi		•	RDF, OWL	●	○	Texte, SparQL, requêtes avancées	●	Jena	RDFstore + DB	Séparées	Wikitext et triplets	●	●	Auto-completion	●	●	Ajax, Java	BSD	stable
Knoodl		•	Unknown	●	●	Texte, SparQL, requêtes avancées	●	○	RDFstore	Dans le texte	wikitext	●	○	○	●	●	Java	Inconnu	stable
KnowWE		•	OWL, d3web XML	●	d3web format	Texte, SparQL	○	Sesame	○	Dans le texte	wikitext	○	●	Auto-completion	●	●	Java	LGPL	beta
OntoWiki		•	RDF(S), OWL	●	○	Texte, SparQL, requêtes avancées	●	pOWL	RDFstore	Séparées	Triplets	●	●	Auto-completion	○	●	PHP	GPL	stable
Semantic MediaWiki	•		RDF(S), OWL	●	○	Texte, SparQL, requêtes avancées	○	○	DB	Dans le texte	wikitext	○	○	○	●	●	PHP	GPL	stable
SMW+	•		RDF(S), OWL	●	○	Texte, SparQL, requêtes avancées	●	KAON 2	DB, RDFstore	Dans le texte	wikitext	●	●	○	●	●	PHP	GPL	stable
SWim		•	RDF(S), OWL	●	OMDoc format	Texte, SparQL, requêtes avancées	○	○	DB	Séparées	wikitext	●	●	○	●	●	Java, XSLT	GPL	alpha
TaOPis	•		F-Logic	●	○	Texte, Flora-2, advanced queries	○	Flora-2	DB	Dans le texte	wikitext	○	○	○	●	●	Python, PHP	GPL	alpha

TABLE 2.1 – Récapitulatif des moteurs de wiki et de leurs fonctionnalités.

également que les principales fonctions attendues d'un wiki sont remplies par la plupart des systèmes et que l'implémentation des standards du Web sémantique est en progression. Toutefois, l'import d'ontologies n'est proposé que par Knoodl, dont les possibités exactes ne sont pas détaillées publiquement. On peut toutefois regretter que certains systèmes intéressants comme AceWiki ou TaOPis souffrent du caractère expérimental de leur développement dans ce type de comparaison.

2.4 Conclusion partielle

À leur création au milieu des années 90, les wikis sont apparus comme des solutions originales pour l'édition collaborative de documents. Leur facilité d'utilisation et leur flexibilité ont conquis de nombreuses communautés d'utilisateurs, à l'image de Wikipédia, mais aussi de wikis plus restreints dans l'industrie. Dans le contexte de l'essor des technologies du Web sémantique, deux approches contribuant à leur amélioration sont apparues. La première propose de se servir de ces technologies pour améliorer les contenus du wiki en y ajoutant des annotations de manière formelle. La seconde consiste à utiliser la philosophie des wikis pour créer des éditeurs d'ontologies collaboratifs en ligne. Pour chacune de ces approches, de nombreux systèmes ont vu le jour, chacun proposant une nouvelle expérience d'utilisation des wikis sémantiques.

Dans la littérature, de nombreuses publications évoquent des expériences de création de wikis sémantiques. Si elles permettent de constater l'utilisabilité des systèmes, peu évoquent les difficultés de maintenance ainsi que la création d'un wiki sémantique à partir d'un contenu existant non structuré. Ce dernier point est le sujet de notre prochain chapitre.

Chapitre 3

OncoLogiK, un wiki sémantique pour les référentiels en oncologie

Sommaire

Pour améliorer le processus d'édition des GBPC, Oncolor a besoin d'un système collaboratif performant qui soit assez simple pour être utilisé par des experts médicaux sans formation spéciale en informatique. Du point de vue du projet KASIMIR, l'ajout d'une couche sémantique permettrait de formaliser, au moins en partie, le contenu des guides afin de proposer des services supplémentaires aux utilisateurs. Au final, la solution retenue est celle du wiki sémantique, répondant aux différents besoins des partenaires du projet.

Dans ce chapitre, la construction de ce wiki sémantique, nommé OncoLogiK, est présentée. La section 3.1 justifie le choix d'un wiki sémantique et met en avant les contraintes d'utilisation d'un tel système dans le contexte médical, ainsi que les décisions à prendre avant de mettre en place le système. Dans la section 3.2, la migration des GBPC existants depuis le précédent site Web d'Oncolor vers une solution collaborative est présentée. Cette migration a donné lieu à une publication en conférence internationale [Meilender *et al.*, 2012b]. Les sections 3.3 et 3.4 évoquent les aspects sémantiques d'OncoLogiK en détaillant respectivement comment les annotations sémantiques sont ajoutées et comment elles sont exploitées. Enfin, la section 3.5 propose une première évaluation de l'outil en recueillant des témoignages d'utilisateurs. Le résultat final a été présenté par Meilender *et al.* [2012a].

3.1 Un wiki sémantique en médecine

3.1.1 Intérêts et contraintes

Pourquoi un wiki sémantique ? Dans ce projet, le choix d'un wiki sémantique comme solution d'édition des GBPC peut être justifié selon plusieurs aspects.

Pour sa part, Oncolor a exprimé le besoin d'un outil collaboratif afin de simplifier le processus d'édition et de permettre une meilleure intégration des experts. De plus, un tel outil permettrait également de récolter des avis externes d'utilisateurs éclairés. Toutefois, pour toucher le plus grand nombre, cet outil doit être facile d'utilisation, étant donné que la plupart des collaborateurs intervenant dans l'édition des GBPC ne sont pas des experts en nouvelles technologies et disposent de peu de temps disponible pour des formations. Dans ce sens, les wikis font figure de solution acceptable pour Oncolor. En effet, ils disposent des fonctionnalités de travail collaboratif nécessaires. De plus, grâce à l'exemple de Wikipédia, la plupart des membres du projet sont confiants dans la facilité d'utilisation de ces systèmes.

Du point de vue du projet KASIMIR, les GBPC sont vus comme des documents contenant des connaissances médicales sous forme textuelle ou graphique. Le but de l'outil est de fournir les procédés qui permettront la formalisation progressive de ces connaissances. En ce sens, les GBPC s'inscrivent dans la notion de continuum de formalisation des connaissances proposée par Baumeister *et al.* [2011a]. Cette notion, qui n'est ni une méthode de formalisation des connaissances ni un modèle de représentation, propose de considérer l'ensemble des connaissances d'un domaine décrites dans différents formats comme une continuité de représentations allant de formes libres (comme les textes ou les images) à des représentations plus formelles (comme des règles logiques). Dans ce travail, les auteurs démontrent qu'un wiki sémantique est un média idéal pour accueillir ce continuum, en particulier parce qu'il permet de conserver plusieurs niveaux de formalisation.

Contraintes en informatique médicale. En plus des contraintes et des limitations des wikis et des wikis sémantiques évoquées dans le chapitre précédent, certains aspects liés au domaine

médical et au processus d'édition sont à prendre en compte. Ainsi, dans une étude sur l'utilisation des wikis pour la création de documents révisables en médecine, Erinoff [2011] liste plusieurs points à éclaircir pour valider leur utilisation.

La notion de statut du document est évoquée : un document n'apparaît pas directement dans sa version finale mais nécessite un long travail collaboratif avant de pouvoir être publié. *A priori*, toutes les pages d'un wiki ne sont donc pas destinées à être rendues publiques dès leur création ou pendant leur édition. Elles nécessitent d'abord des ajouts, des validations ou des corrections. Or, les GBPC sont des données médicales sensibles et, en particulier dans le cadre de l'application de la charte HONcode (présentée dans la section 1.2.2.2), de tels documents ne peuvent pas être diffusés tels quels. Afin que la confusion ne soit pas possible, on peut établir une différenciation entre les notions de document final et de brouillon. La première représente le document dans sa dernière version diffusable tandis que la seconde correspond à la version la plus récente dont le contenu n'a pas encore été validé. Dans un système d'information, cette différenciation se traduit par une distinction des niveaux d'accessibilité et, de préférence, par une identification visuelle.

Le problème de gouvernance interne du système d'information est également évoqué. Il convient de déterminer de manière claire quels sont les accès autorisés aux différents profils d'utilisateurs. Par exemple, il semble important pour des raisons de crédibilité de ne pas laisser des documents à l'état de brouillon en accès publique. De même, il faut déterminer quelles seront les autorités à même de trancher dans un conflit d'édition, c'est-à-dire à même de choisir quelle est la version à diffuser parmi plusieurs propositions. Cette autorité doit également être en charge de la responsabilité de la diffusion des informations. Ainsi, elle décide à quel moment un brouillon peut être diffusé, et donc passer au statut de document final.

3.1.2 Intégrer les GBPC dans un wiki sémantique

***Wikis for Ontologies* ou *Ontologies for Wikis* ?** Avant d'installer le wiki, il convient de décider quelle approche des wikis sémantiques implémenter parmi celles présentées dans la section 2.2.2. L'approche OfW contraint fortement l'édition. Elle impose une ontologie préalablement définie, qui soit de préférence alignée avec d'autres ressources ontologiques du domaine. À l'opposé, l'approche WfO accorde plus de libertés aux éditeurs. Toutefois, l'absence d'une structure commune restreint les possibilités d'ajout de services sémantiques. En effet, si les documents présentent des annotations disparates, il sera impossible de les comparer ou de les classifier entre eux et, *a fortiori*, de les confronter à des ressources externes.

Dans le cas de notre application, il n'existait aucune ontologie préalable. Du point de vue d'Oncolor, il est important pour les utilisateurs de mettre en avant le contenu des GBPC plutôt que les annotations syntaxiques. En effet, la plupart des utilisateurs finaux s'intéressent uniquement à ces contenus et ne trouvent d'intérêt pour les annotations sémantiques qu'à travers les services qu'elles permettent. Il convient donc que ces annotations soient les plus transparentes possible afin de ne pas perturber la navigation. De plus, comme évoqué dans la section 1.2.2, les GBPC sont rédigés par des experts médicaux dont les compétences en ingénierie des connaissances sont limitées. Il faut donc encadrer le plus fortement possible la création des annotations afin de conserver leur cohésion, sans qu'elle ne gène ni la rédaction ni la navigation.

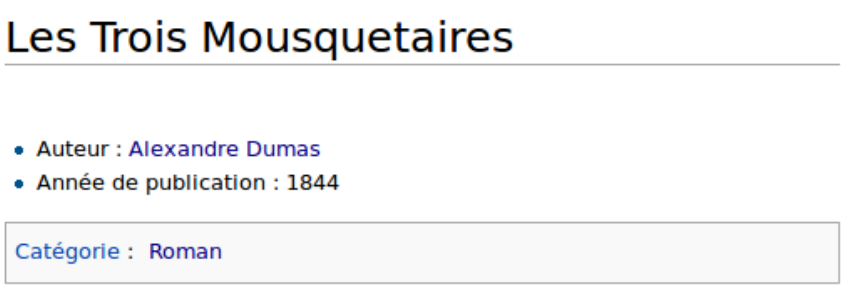
Une solution alternative revenant à combiner les deux approches peut être étudiée. Elle consiste à utiliser une approche WfO en y ajoutant la notion de modèles (*template*) [Haake *et al.*, 2005]. Dans le cadre des wikis, les modèles sont des éléments structurants, aussi bien pour l'apparence que pour le contenu. Ils permettent de définir des séquences paramétrables qui peuvent être appliquées sur plusieurs pages. Ainsi, les modèles peuvent être utilisés pour définir l'usage des annotations sémantiques. L'application systématique de tels modèles aux différents

```
* Auteur : [[aPourAuteur::{{{auteur}}}]]
* Année de publication : [[aPourAnnée::{{{année}}}]]
[[Catégorie:Roman]]
```

(a) Définition du modèle `Modèle:Roman`.

```
{{Roman
|auteur=Alexandre Dumas
|année=1844
}}
```

(b) Appel du modèle `Modèle:Roman` dans la page `Les Trois Mousquetaires`.



(c) Affichage de la page `Les Trois Mousquetaires`.

FIGURE 3.1 – Exemple d'utilisation d'un modèle dans SMW.

GBPC permettra de conserver la cohésion de la structure tout en limitant leur visibilité pour les utilisateurs et les rédacteurs.

Un exemple de modèle et son application à une page dans le moteur SMW sont donnés dans la figure 3.1. Dans ce cas, le but du modèle est de systématiser l'emploi des mêmes propriétés et des mêmes catégories pour un type de page. Ainsi, l'emploi de ce modèle dans chaque page de roman du wiki en adaptant les paramètres garantit que chacune de ces pages utilisera les propriétés `aPourAuteur` et `aPourAnnée` et la catégorie `Roman`.

Implémenter l'approche documentaire. La section 1.2.3 présente les deux approches pour l'informatisation des GBPC : l'approche documentaire et l'approche centrée sur les connaissances. L'intégration des GBPC dans un système collaboratif sémantique relève de l'approche documentaire. En effet, le but est de reprendre les guides diffusés sur le site d'Oncolor pour y adjoindre de nouveaux services. Ces services seront de deux natures :

- intégration d'outils collaboratifs du Web 2.0 pour faciliter la rédaction et la communication entre professionnels,
- ajout d'annotations et, plus globalement, intégration des technologies du Web sémantique dans le but d'améliorer l'utilisation des GBPC en ligne.

Toutefois, l'approche centrée sur les connaissances n'est pas ignorée dans le projet. Le but de cette approche est d'obtenir des GBPI entièrement formalisés pouvant être utilisés par des processus automatiques. Ainsi, elle peut être vue comme un but à long terme pour le projet KASIMIR. L'intégration dans le processus d'édition d'outils sémantiques peut constituer une première étape vers une formalisation plus complète des GBPC.

Approches précédentes. Les wikis sémantiques sont déjà utilisés dans la communauté médicale. Par exemple, Papakonstantinou *et al.* [2011] en proposent un pour l'entraînement à la prescription en ligne. Le wiki repose sur une ontologie dont le but est de vérifier la cohérence de la prescription par rapport à des connaissances médicales et des connaissances sur le patient. Pour leur part, He *et al.* [2009] exploitent la simplicité des annotations sémantiques pour la création collaborative d'une terminologie médicale. Les fonctions sémantiques sont utilisées par Jiang *et al.* [2010] pour trouver des points communs et harmoniser la description des données dans les études cliniques. Tous ces exemples ont en commun de proposer des wikis qui sont soit complètement ouverts, soit complètement fermés, et qui ne sont pas créés à destination du grand public. Aucun ne pose la question du statut des documents brouillons.

Dans la littérature, seulement deux approches pour l'informatisation des GBPC reposant sur les wikis sémantiques ont été proposées. La première approche, nommée CliP-MoKi, est présentée par Rakebul [2009]. Le but est de proposer un environnement collaboratif pour la création de GPBI au format Asbru. Ce format est un exemple de langage de GBPI de l'approche centrée sur les connaissances. Son utilisation nécessite donc une reprise totale et une traduction des GBPC dans ce format. D'un point de vue technique, CliP-MoKi utilise le moteur de wiki sémantique SMW en y intégrant des formulaires adaptés et des extensions spécifiques.

L'autre approche [Hatko *et al.*, 2012] repose sur le moteur KnowWE, présenté dans le chapitre précédent. Fondé sur le principe du continuum de formalisation des connaissances, elle propose un environnement collaboratif assez complet pour l'édition des GPBC. Toutefois, la solution ne dispose d'aucun lien vers des ontologies médicales connues et repose sur l'utilisation de formats propriétaires tels que *d3web*.

3.2 Migration d'un site statique vers un wiki

La section précédente a mis en avant les choix théoriques qui guident la création du système. Du point de vue des wikis sémantiques, l'approche choisie est l'approche WfO, complétée par l'utilisation de modèles. Du point de vue de l'informatisation des GBPC, notre approche s'intègre dans l'approche documentaire en se concentrant particulièrement sur l'ajout de services à destination des utilisateurs.

Le but de cette section est de présenter la manière dont les GBPC existants sur l'ancien site Web d'Oncolor ont été importés en implémentant ces principes.

La section 3.2.1 justifie l'utilisation de SMW comme moteur de wiki sémantique. L'import des GBPC dans ce moteur est présenté dans la section 3.2.2. L'emploi de ce nouvel outil dans le processus d'édition nécessite la définition d'une nouvelle politique éditoriale : celle-ci est présentée dans la section 3.2.3. Enfin, le but de la section 3.2.4 est de mettre en avant les éléments de cette migration généralisable à d'autres migrations de ce type.

3.2.1 Choix du moteur de wiki

Pour répondre aux exigences des différents acteurs du projet Kasimir, le choix du moteur dépend de plusieurs critères scientifiques et techniques, au niveau de ses capacités, de son installation, de son utilisation et de sa maintenance. Une liste de ces critères peut être établie :

- Le système doit être dans une version stable. Le caractère industriel de son implémentation contraint à écarter les systèmes uniquement disponibles dans des versions alpha ou bêta, qui comportent par nature encore des défauts de programmation et potentiellement des failles.

- Il doit également avoir déjà été éprouvé à large échelle, ce qui garantit sa qualité.
- Son développement doit être *open source*, encadré par une large communauté ce qui favorise la présence d'extensions.
- L'édition doit être simple d'utilisation, abordable pour des experts médicaux non spécialistes des wikis sémantiques.
- Il doit disposer de toutes les fonctionnalités nécessaires à la mise en œuvre d'une politique éditoriale compatible avec les préceptes d'édition des GBPC et proposer des possibilités d'échange entre les utilisateurs.
- Les annotations sémantiques doivent pouvoir se faire de deux manières : de manière libre, c'est-à-dire que les utilisateurs peuvent ajouter une annotation à partir de propriétés dont ils définissent eux-même la sémantique, et de manière contrainte, c'est-à-dire selon les propriétés créées par les administrateurs du système. L'annotation contrainte peut être facilitée par l'usage de modèles.
- Enfin, il doit proposer des possibilités d'ouverture pour les technologies du Web sémantique, en intégrant par exemple un *RDFstore* ou un point d'accès SPARQL.

Après confrontation de ces exigences à l'état de l'art, la solution composée de MW et de son extension SMW a été retenue. La plupart des autres moteurs manquent d'exemples de déploiement à grande échelle et reposent sur des communautés de développeurs réduites.

3.2.2 Import des GBPC

Installation et configuration du moteur. L'installation et la configuration de MW et de son extension SMW sont assez intuitives et largement documentées sur leurs sites Web respectifs. Ces sites proposent également de nombreuses extensions complémentaires dont plusieurs ont été testées. Certaines ont été installées et parfois modifiées pour correspondre aux besoins spécifiques de l'application.

Intégration des contenus. Une fois le wiki installé sur le serveur, l'étape suivante a consisté à importer les GBPC depuis l'ancien site d'Oncolor. Le processus de migration est illustré par la figure 3.2 qui montre les étapes réalisées. L'ensemble des tâches réalisées de manière automatique lors de cette migration a donné lieu à la création d'un *parser* dédié.

La première étape de la migration a consisté à faire l'état des lieux des ressources existantes. Le site diffusait alors 146 GBPC, chacun étant composé de une ou plusieurs pages parfois liées à des images ou des fichiers PDF. Il a donc fallu rassembler pour chaque guide l'ensemble de ses composants, en respectant la chronologie du document, disponible à partir du menu de la page d'accueil du référentiel. Ensuite, dans chaque page HTML, un « nettoyage » a été effectué afin d'isoler le contenu des GBPC des éléments de navigation. Le code de ces pages est le résultat de mises à jour successives des pages avec différents éditeurs WYSIWYG utilisant de différentes manières différents langages Web (HTML, javascript, CSS, XHTML) dans différentes versions de ces langages. En définitive, la plupart des pages ne peuvent pas être considérées comme syntaxiquement correctes dans l'ensemble de ces langages. Cet état de fait excluait donc d'utiliser la plupart des *parsers* XML classiques.

Les parties suivantes visent à identifier les champs de texte présents dans la majorité des référentiels dont le contenu serait intéressant à formaliser dans le wiki sémantique. Ces champs peuvent être de plusieurs natures : ils peuvent contenir la date de mise à jour, des mots-clés, etc. Pour recevoir ces textes, des modèles sont édités dans le wiki. Pour chaque champ, une

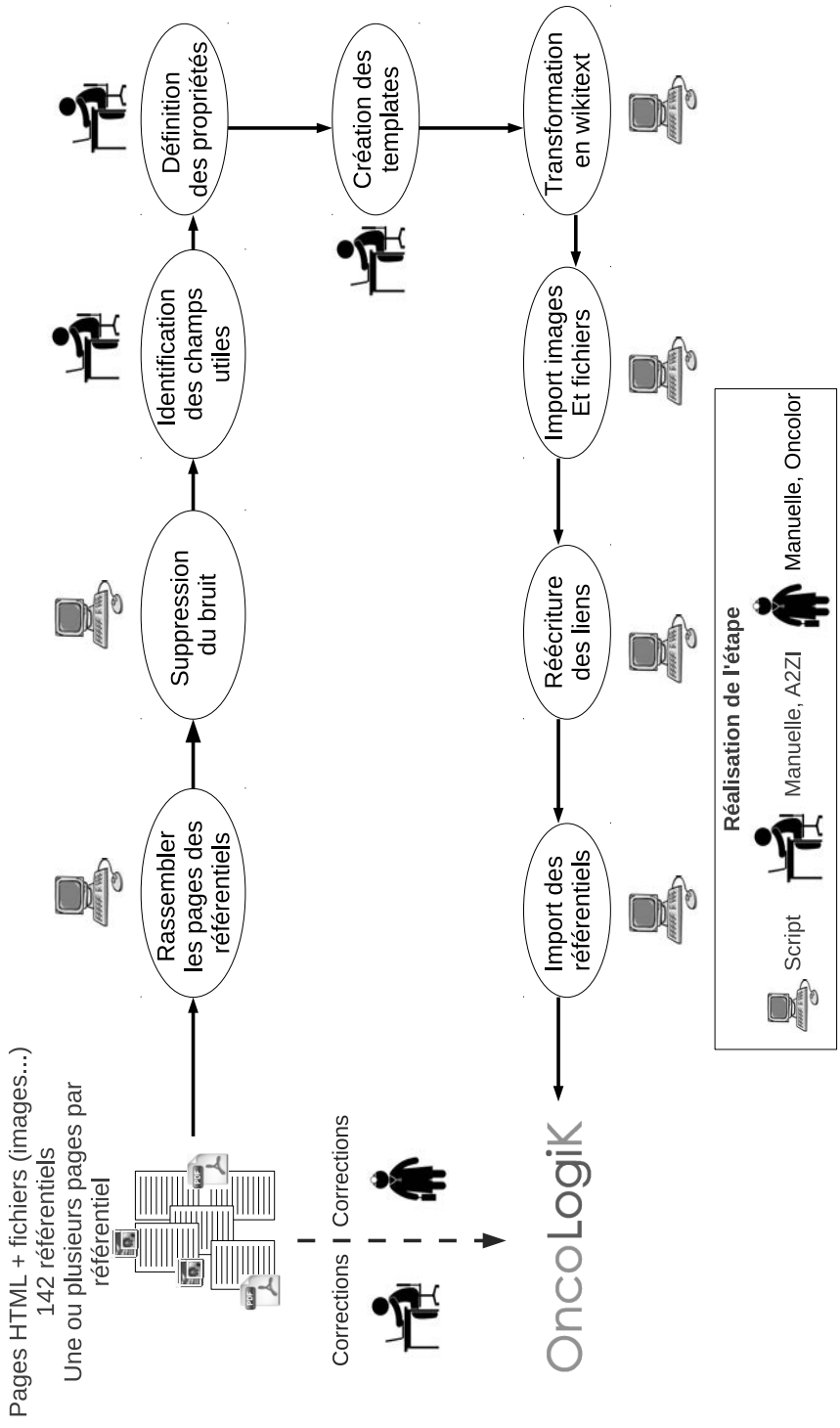


FIGURE 3.2 – Représentation du processus de migration vers un wiki sémantique.

propriété est créée afin d'exprimer le sens de son contenu. Le *parser* est adapté pour isoler ces champs et les intégrer dans les modèles d'OncoLogiK. On peut noter qu'une utilisation correcte des standards HTML et CSS aurait permis de distinguer bien plus simplement le contenu des éléments de présentation et de repérer plus facilement ces champs spéciaux.

La suite du processus porte sur l'intégration même du contenu des GBPC dans le wiki. Elle repose largement sur les fonctions d'import disponibles dans le *framework* de MW. Le contenu des pages des GBPC est transformé en wikitext, en s'appuyant sur le code HTML et les styles CSS. En particulier, le *parser* transforme les structures HTML et CSS simples de formatage de texte (du type couleur, italique, etc.) en wikitext. Il fait de même avec des structures plus complexes du type liste ou tableau. Un algorithme a également été ajouté afin de détecter et de formater les éléments des plans des GBGC (titre et sous-titres sur 4 niveaux).

Les fichiers annexes aux GBPC (images, PDF, etc.) sont importés, et les liens des GBPC vers ces ressources sont adaptés à leur nouvelle adresse dans le wiki. Les GPBC ainsi formatés sont finalement importés, et les liens entre eux sont corrigés.

L'état des codes sources des GPBC n'a cependant pas permis de fournir des pages dans leur version définitive. Un large travail de reformatage et de correction a donc été nécessaire. En effet, de nombreuses structures n'ont pas été reconnues, aussi bien le formatage du texte que des champs et des sous-titres. Les référentiels ont enfin été validés par les membres d'Oncolor pour permettre l'ouverture du wiki au public. Ces délais de transition ont donné lieu à un surplus de travail pour les rédacteurs puisque, durant cette période, les modifications ont été faites en double : sur l'ancien site et sur le wiki. Le point positif a été de montrer aux utilisateurs durant cette transition les avantages de l'édition sur les wikis et la présence des fonctionnalités nécessaires au processus d'édition. Cette période peut ainsi être vue comme une période de formation à l'outil.

En parallèle à l'import des GBPC, un thésaurus de pharmacologie également édité par Oncolor a été intégré dans le wiki. Son contenu est fortement lié au contenu d'OncoLogiK. Il est destiné à recenser des informations utiles pour la préparation de produits pharmaceutiques évoqués dans les parties concernant le traitement de certains GBPC. Concrètement, une page est consacrée à chaque produit présenté. La taille de ce thésaurus étant relativement réduite (une vingtaine de pages assez courtes par rapport au GBPC), son import a été manuel.

3.2.3 Politique éditoriale d'OncoLogiK

3.2.3.1 Niveaux d'utilisateurs et les espaces d'édition

Identifier et créer les profils d'utilisateurs. Comme évoqué précédemment, un wiki médical ne peut être une application complètement ouverte. Afin de créer les profils d'utilisateur du site et d'établir les droits correspondants, les contraintes d'utilisation de wiki en médecine, le processus d'édition des GBPC et l'observation de l'utilisation du site précédent ont été pris en compte. Ainsi, cinq types d'utilisateurs ont été identifiés :

- Les visiteurs réguliers ou occasionnels consultent les référentiels sans y contribuer. Ce sont généralement des patients, des médecins généralistes ou des étudiants en médecine.
- Les visiteurs « éclairés » sont des professionnels en médecine, non membres d'Oncolor. S'ils ne contribuent pas directement aux contenus, ils peuvent commenter, proposer des corrections ou mettre en avant des manques. Ces contributions doivent être discutées par les membres d'Oncolor avant de donner lieu à des modifications.

- Le personnel médical d'Oncolor réunit les principaux contributeurs des GBPC.
- Le personnel technique d'Oncolor propose un support pour faciliter et améliorer les GBPC en assurant par exemple la mise en page.
- Les administrateurs fournissent le support technique informatique.

À partir de ces profils, des groupes d'utilisateurs sont définis et paramétrés dans OncoLogiK, chacun ayant des droits correspondants à ses fonctions :

- le groupe « *utilisateur* » représente des utilisateurs anonymes non connectés n'ayant que des droits de lecture dans les zones publiques ;
- le groupe « *lecteur* », correspondant aux visiteurs « éclairés », est un ensemble d'utilisateurs connectés avec accès en lecture et un droit de commenter les référentiels ;
- le groupe « *rédacteur* » correspond au staff médical qui a accès en écriture aux référentiels ;
- les membres du groupe « *oncolor* » (staff technique d'Oncolor) ont accès à toutes les pages, avec des fonctions d'administration limitées ;
- les membres du groupe « *administrateur* » sont autorisés à faire toutes les modifications possibles.

Intégration des espaces d'édition. Dans la section 3.1.1 est posé le problème du statut du document dans un wiki médical. Dans le cadre d'OncoLogiK, le processus d'édition et de certification impose plusieurs contraintes, dont celle de ne publier que les référentiels dont le contenu a été validé. Les référentiels en cours d'élaboration ou de modification ne doivent donc pas être accessibles au public mais uniquement aux utilisateurs concernés par leur élaboration.

Afin d'affiner les possibilités de contrôle d'accès, MW propose un système de gestion des espaces d'édition. Un espace peut être défini par un nom et les pages qui le composent sont soumises à des droits d'édition spécifiques à cet espace. On peut d'ailleurs noter que MW utilise ce système par défaut pour isoler par exemple les « pages spéciales » qui contiennent les fonctions d'administration ou les pages gérant l'intégration de fichiers multimédias. Dans le cadre de SMW, pour chaque espace, un espace de nom spécial est créé et est utilisé par toutes les pages de cet espace. Ainsi, pour coller aux contraintes d'édition des GBPC, un espace de travail peut être créé. Chaque page de cet espace représente le brouillon d'un référentiel, c'est-à-dire une version sujette à des modifications qui n'ont pas encore été validées.

Ainsi, plusieurs espaces ont été créés :

- l'espace public : espace sans nom spécial, destiné à contenir les référentiels en version définitive, accessible en lecture à tous les utilisateurs mais accessible en écriture seulement aux personnes autorisées.
- l'espace `Travail` : espace de travail, où se trouvent les référentiels en cours de mise à jour.
- l'espace `Arbre` : espace destiné à recevoir les arbres de décision édités par KCATOS qui est présenté dans le chapitre 5.
- l'espace `Pharma` : espace pour le stockage du thésaurus de pharmacologie.
- l'espace `Mesh` : espace pour le stockage des termes MeSH dédiés à l'indexation des référentiels.
- l'espace `Maintenance` : espace pour les pages dédiées à la maintenance du site et à l'aide en ligne aux utilisateurs.
- l'espace `Rédacteur` : destiné à recevoir les listes de personnes intervenant dans la rédaction pour chaque GBPC.

Ces espaces sont présentés dans la table 3.1 qui indique également les droits d'édition qui lui sont associés.

	Admin	Oncolor	Rédacteur	Lecteur	Anonyme
Arbre	Écriture, lecture, discussion	Écriture, lecture, discussion	Écriture, lecture, discussion	Aucun	Aucun
Fichier	Écriture, lecture, discussion	Écriture, lecture, discussion	Lecture, discussion	Lecture, discussion	Lecture
Maintenance	Écriture, lecture, discussion	Écriture, lecture, discussion	Lecture, discussion	Lecture, discussion	Lecture
Mesh	Écriture, lecture, discussion	Écriture, lecture, discussion	Lecture, discussion	Lecture, discussion	Lecture
Pharma	Écriture, lecture, discussion	Écriture, lecture, discussion	Écriture, lecture, discussion	Lecture, discussion	Lecture
Public	Écriture, lecture	Lecture	Lecture	Lecture	Lecture
Travail	Écriture, lecture, discussion	Écriture, lecture, discussion	Écriture, lecture, discussion	Lecture, discussion	Aucun

TABLE 3.1 – Espaces d'édition et droits associés dans OncoLogiK.

Processus de mise à jour de GBPC. Après la validation d'un GBPC, sa mise à jour consiste donc à le transférer de l'espace de travail à l'espace public, en écrasant la version précédente. Toutefois, un référentiel n'est pas uniquement composé de la seule page de wiki : il peut également comporter des images, des arbres de décision ou des fichiers multimédias stockés dans d'autres pages. Pour que la version publique soit cohérente, il faut donc conserver une trace de ces dépendances au moment de la publication, de manière à ce que la version liée à la page ne soit pas une version modifiée non validée.

Pour adapter MW à cette spécificité, une extension a été implantée dont le principe est de créer une version spéciale des dépendances lors de la mise à jour. Ainsi, le référentiel en accès public est lié à cette version spéciale de l'image ou du document qui n'est pas modifiable par les rédacteurs. Ce processus peut être illustré par la figure 3.3. D'un point de vue utilisateur, la mise à jour se fait grâce à l'utilisation d'une page spéciale, accessible à l'équipe technique d'Oncolor et aux administrateurs.

3.2.3.2 Favoriser l'acceptation de l'outil par les utilisateurs

Comme la plupart des systèmes d'information, un wiki peut sembler de prime abord complexe d'utilisation pour un non expert. Afin de favoriser l'acceptation et donc l'utilisation d'OncoLogiK, il est nécessaire de montrer sa simplicité d'utilisation. Mader [2007] définit la notion de *wikipatterns*. Le but de ces *patterns* est de proposer un ensemble de modèles d'utilisation favorisant l'utilisation de l'outil. Ils sont construits empiriquement afin de transmettre les retours d'expérience. *A contrario*, des *anti-patterns* sont également définis dans le but d'identifier les techniques et les comportements contre-productifs. Un site Web collaboratif [Wikipatterns.com, 2013] a ainsi été créé dans le but de recenser les *wikipatterns*. À titre d'exemple, la gestion des droits d'utilisateur en fonction des espaces peut être vue comme l'adaptation d'un *pattern* dont le but est de simplifier l'utilisation du wiki en préconisant de créer des espaces selon les droits d'accès aux documents.

Parmi les initiatives destinées à favoriser l'utilisation d'OncoLogiK, on peut citer notamment :

– la création d'une documentation technique à destination des administrateurs [Meilender,

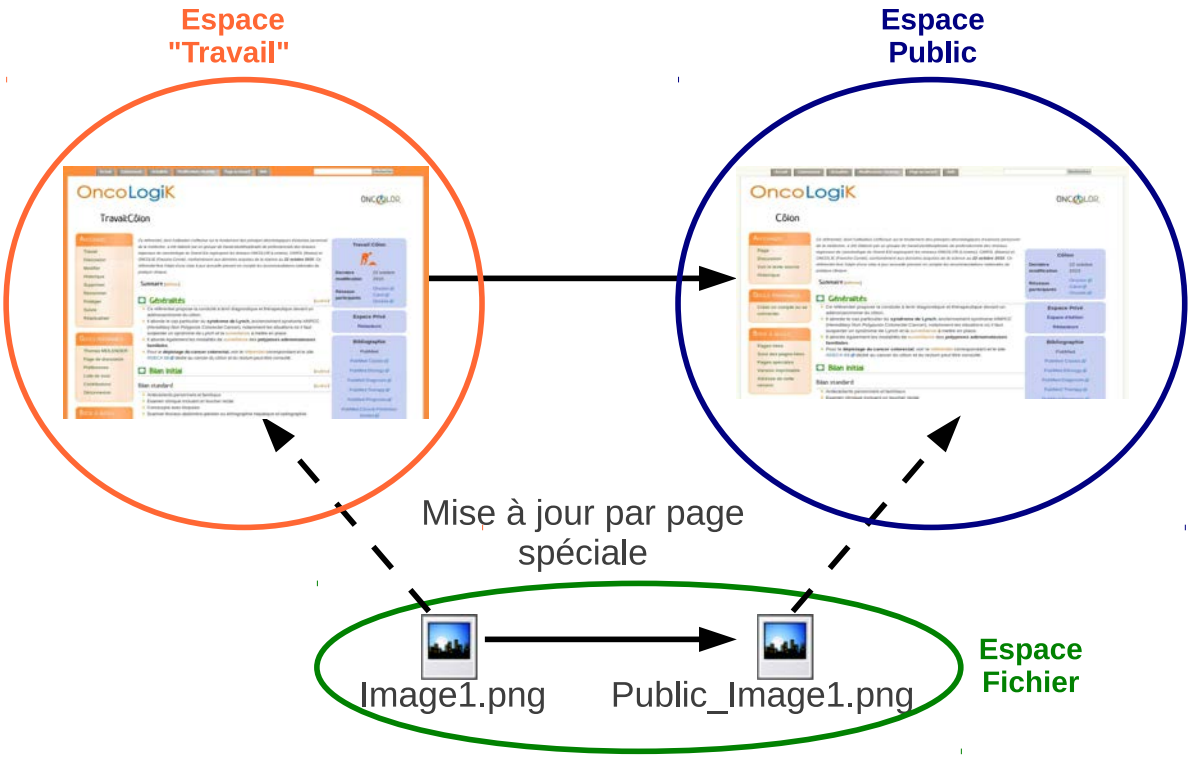


FIGURE 3.3 – Processus de mise à jour des GBPC.

2012b] ;
- l'animation de pages de discussion et de séminaires pour l'aide au staff technique d'Oncolor ;
- la création de notices d'utilisation à destination de la communauté dans l'espace `Maintenance` sur le wikitext et l'édition d'arbres de décision ;
- la création d'un mode « bac à sable », c'est-à-dire un espace permettant de faire des essais de rédaction de manière libre, sans que les modifications ne soient considérées comme des publications, pour permettre une utilisation libre de l'outil afin de se former.

3.2.4 Vers des éléments méthodologiques pour la migration

En analysant cette expérience de migration d'un site Web statique existant depuis plusieurs années à une solution du Web social sémantique dans le domaine médical, il apparaît que plusieurs points peuvent être améliorés et que certaines étapes pourraient être généralisées pour des migrations concernant d'autres domaines. De cette réflexion est née une proposition de méthodologie permettant de guider ce type de migration. La première version de cette méthode est présentée dans la figure 3.4. Elle est appelée à être testée sur d'autres projets afin de pouvoir être corrigée, complétée et validée.

Cette méthode repose sur la connaissance du contexte d'édition, c'est-à-dire l'analyse du processus d'édition et des contenus. À partir de ces éléments peut être choisi le système d'information adéquat. Selon le profil des personnes impliquées dans le processus, ce système peut être plus ou moins complexe à utiliser ou nécessiter des outils de communication spécifiques. Pour sa

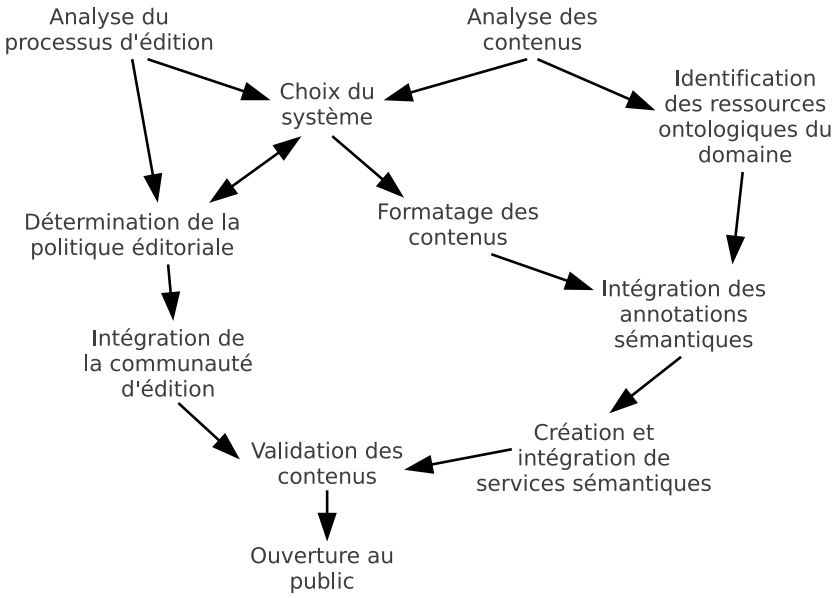


FIGURE 3.4 – Éléments méthodologiques d'une migration d'un site Web statique vers une solution du Web social sémantique.

part, l'analyse des contenus peut fournir une idée sur le type d'outils d'édition à déployer.

En fonction du système, les contenus peuvent être repris, formatés de manière à correspondre à ce système et importés. Il est également possible d'y intégrer des annotations sémantiques. Afin de pouvoir exploiter les capacités du Web sémantique, il est intéressant de consulter les ressources ontologiques déjà existantes. Ces ressources peuvent ainsi fournir un langage commun avec d'autres systèmes sémantiques. En fonction des besoins des utilisateurs et du type d'annotation choisi, des services sémantiques peuvent être proposés afin de fournir une plus-value à l'outil.

Le choix du système et l'analyse du processus d'édition permettent d'élaborer la politique éditoriale. Elle se traduit par la création des groupes d'utilisateurs et la définition de leurs droits, ainsi que la création d'espaces d'édition. Cette politique éditoriale peut également influer sur le choix du système dans le sens où une politique trop spécifique peut ne pas être applicable dans certains systèmes. En fonction des droits attribués, les principaux éditeurs peuvent être formés à l'utilisation du système. Pour faciliter cette formation, des initiatives peuvent être prises : dans notre application, l'adoption de *wikipatterns* en est un exemple.

Un des avantages de bien former ces éditeurs est de pouvoir disposer de relais intégrés à la communauté lors de l'ouverture au public. Enfin, avant d'ouvrir le système, il est important de valider les contenus repris et les services sémantiques proposés. Cette activité constitue un très bon complément à la formation des éditeurs qui peuvent mettre en pratique leur connaissance de l'outil.

3.3 Annotation des GBPC

3.3.1 Catégorisation

Dans le cadre des wikis, la catégorisation des pages est largement utilisée, en particulier pour faciliter la navigation. Les wikis sémantiques en tirent parti en considérant les pages comme les classes de l'ontologie sous-jacente. Dans OncoLogiK, le premier niveau de catégorisation est extrait du site Web précédent. Dans ce dernier, un travail de classification des référentiels avait été réalisé. Cette classification reposait sur des catégories médicales créées pour l'application, assez proches de concepts présents dans des classifications déjà existantes. Ainsi, les GBPC étaient classés selon 16 axes concernant pour la plupart l'anatomie, comme par exemple `Appareil respiratoire`, `Système hématopoïétique` et `Système nerveux`. Certains de ces axes (comme par exemple la catégorie `Divers`) sont toutefois plus difficiles à aligner avec une classification standard. Afin de faciliter la transition vers le wiki sémantique et de manière à ne pas perturber les habitués du site, ces axes ont été conservés tels quels. On peut également noter que cette catégorisation est en cours d'évolution en se servant du wiki sémantique comme support. Le but est de l'affiner en ajoutant des sous-niveaux pour chaque axe, en conservant une perspective proche des classifications anatomiques.

Pour améliorer la navigation, d'autres axes de catégorisation ont été créés, en concertation avec Oncolor. Ces axes concernent :

- l'éditeur du document (étant donné qu'OncoLogiK héberge désormais également des GBPC édités par d'autres réseaux régionaux),
- son statut, de manière à identifier formellement les brouillons et les documents en version finale,
- le type de document, qui sert à montrer si le document est un référentiel, un arbre de décision, une page du thésaurus de pharmacologie, etc.

L'exploitation de ces axes a donné lieu à la création de la notion de GBPC proche. Un GBPC est considéré comme proche d'un autre s'ils partagent les mêmes catégories, ou d'un point de vue sémantique, s'ils font partie des mêmes classes. L'idée est qu'un utilisateur intéressé par un référentiel sera également intéressé par les référentiels qui lui sont proches. Pour améliorer la navigation, un composant graphique spécial a été créé afin de lister tous les GBPC proches de celui consulté. Son fonctionnement repose sur l'emploi de modèles et de requêtes sémantiques[37].

3.3.2 Annotations guidées

Le principal attrait des wikis sémantiques de l'approche WfO est la liberté laissée à l'utilisateur quant à la création d'annotations sémantiques. OncoLogiK espère d'ailleurs tirer profit de cette liberté en recueillant le plus possible de connaissances fournies par les utilisateurs. Toutefois, l'édition d'annotations nécessite certaines compétences en ingénierie des connaissances, en particulier pour distinguer les connaissances importantes à ajouter. De plus, pour pouvoir utiliser automatiquement ces connaissances, une certaine régularité doit apparaître dans l'édition afin de trouver les mêmes propriétés à chaque page. Les experts médicaux éditant les GBPC ne disposant pas toujours des compétences nécessaires, des modèles ont été créés afin de les guider dans les processus d'annotations. Ces modèles gèrent la saisie, le traitement et l'affichage des informations.

Ces annotations sont considérées comme guidées dans le sens où leur présence est imposée par l'emploi du modèle qui détermine quelles sont les propriétés utilisées. Elles laissent toutefois

37. D'autres applications des requêtes sémantiques sont présentées dans la section 3.4.1

```
{{Infobox Référentiel
|Date modification=16 février 2011
|Date séminaire=22 octobre 2010
|Méthode de mise à jour=Inconnue
|Coordinateur=Inconnu
|Rédacteur=Inconnu
|Réseaux participants1=[http://www.oncolor.org Oncolor]
|Réseaux participants2=[http://www.reseau-carol.fr Carol]
|Réseaux participants3=[http://www.oncolie.fr Oncolie]
|Mot clé1=Mesh:Cancer de l'appendice
|Lien=Aucun
|URL=[http://www.oncolor.org/referentiels/appendice_acc.htm]
}}
```

(a) Appel du modèle `Référentiel`.

(b) Affichage correspondant.

FIGURE 3.5 – Annotations sémantiques d'un référentiel par un modèle et affichage résultant.

toute liberté à l'utilisateur concernant leur contenu, en proposant parfois un formatage spécial pour l'affichage, comme pour les dates par exemple.

La figure 3.5 donne l'exemple d'un modèle permettant la saisie de ces annotations (3.5a) et de l'affichage correspondant (3.5b) pour le GBPC concernant les tumeurs appendiculaires. Dans cet exemple, les annotations permettent de préciser les dates de mise à jour et de séminaires concernant ce référentiel, la méthode de mise à jour utilisée, les personnes responsables de son édition, les réseaux participant à sa rédaction, les mots-clés liés, des liens vers des ressources externes liées et l'URL de la page de ce GBPC sur l'ancien site. Seule une partie de ces informations apparaît lors de l'affichage, selon les volontés d'Oncolor. Les autres sont destinées à un traitement interne.

3.3.3 Annotations avec MeSH

Utiliser un thésaurus médical. OncoLogiK propose également d'utiliser des annotations plus structurées, reposant sur des ressources médicales déjà existantes. Afin de fournir des liens avec des applications externes du Web sémantique, il faut utiliser un vocabulaire compréhensible par d'autres applications. Les RTO médicales présentées dans le chapitre 1 fournissent de tels vocabulaires. On peut constater que les RTO sont généralement le fruit d'une réflexion menée en fonction de l'application qui en est faite. Dans le cadre d'OncoLogiK, les objets modélisés sont des documents textuels destinés à servir de référence pour les praticiens.

Parmi toutes les ressources présentées précédemment, MeSH semble être le vocabulaire le mieux adapté au contexte documentaire. En effet, MeSH est spécialement conçu pour l'indexation de documents médicaux et pour la recherche d'information. Sa structure peut être vue comme une hiérarchie de termes couvrant tous les domaines de la médecine. En outre, il existe une version francophone de ce thésaurus et plusieurs experts d'Oncolor connaissent déjà son principe de représentation pour l'avoir utilisé sur des sites Web de publications médicales tels que PubMed.

Import de MeSH. Ainsi, des termes d'indexation du thésaurus MeSH ont été importés automatiquement dans OncoLogiK. Pour ce faire, un *parser* spécifique a été développé, répondant aux besoins du projet. En effet, il n'est pas nécessaire à ce moment d'importer l'intégralité du

FIGURE 3.6 – Terme MeSH « Tumeurs de l'œsophage » tel qu'importé dans OncoLogiK.

thésaurus. Pour ne pas alourdir inutilement la base de données du wiki, certains axes non nécessaires dans ce contexte ont été ignorés, comme par exemple l'axe concernant la géographie ou celui sur la biologie animale. De plus, pour chaque terme, un nombre conséquent d'informations est donné tels que les synonymes en anglais. Après avoir déterminé les informations importantes à conserver, un modèle a été créé dans le wiki afin de gérer leur intégration. Ce modèle gère des propriétés afin de relier le terme aux informations importées. Au final, chaque terme MeSH importé dans le wiki s'affiche à la manière de celui présenté dans la figure 3.6.

On peut noter que la présence de services sémantiques dans le modèle compense l'absence des informations ignorées lors de l'import des termes. Si ces informations ne sont pas directement dans la page, elles restent accessibles par la génération automatique de liens vers des pages présentant ces termes sur les sites Web PTS et BioPortal.

Utilisation pour l'annotation des référentiels. Afin de faciliter l'annotation des GBPC, un champ a été prévu dans le modèle présenté dans la figure 3.5a. Pour chaque guide, un ou plusieurs terme(s) MeSH peu(ven)t être sélectionné(s). Le choix est laissé à l'utilisateur quant à la manière dont ces termes se coordonnent : ils peuvent être reliés soit par des conjonctions, soit par des disjonctions grâce au paramétrage du modèle. Au final, les termes indexant les référentiels sont visibles par tous les utilisateurs dans un modèle d'affichage spécial en bas de page. De plus, pour chaque terme MeSH, une requête est ajoutée dans leur modèle de manière à recenser tous les GBPC l'utilisant. Concrètement, l'annotation des GBPC dans OncoLogiK a été réalisée par des experts médicaux d'Oncolor, ce qui a donné lieu à un important travail de documentation et d'indexation.

3.4 Exploiter les annotations sémantiques

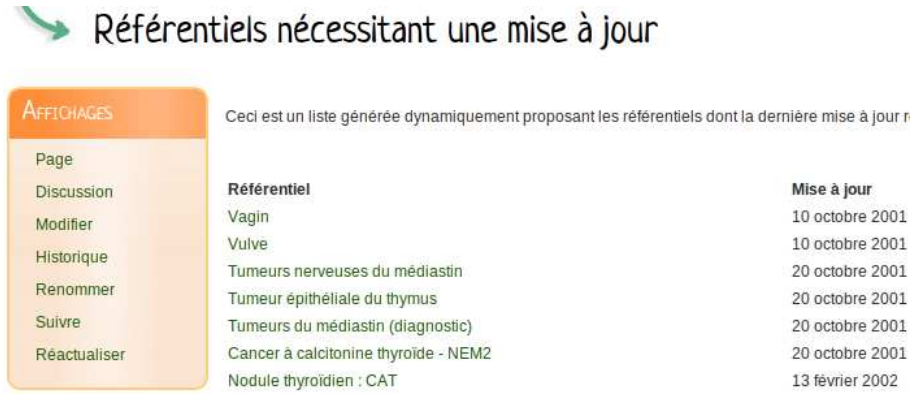
3.4.1 Le moteur de requêtes de SMW

SMW propose un langage simple de requêtes sémantiques [38]. Ces requêtes peuvent être in-

38. `http://semantic-mediawiki.org/wiki/Help:Inline_queries`

```
{{#ask:[[Catégorie:Référentiel]] [[a pour date de mise à jour::<{{#time:d F Y|2 years ago}}]]
|?a pour date de mise à jour
| sort=a pour date de mise à jour
| format=template
| template=table milieu
| introtemplate=table haut
| outrotemplate=table bas
}}
```

(a) Exemple de syntaxe d'une requête sémantique dans SMW.



(b) Résultat de la requête précédente dans OncoLogiK.

FIGURE 3.7 – Exemples de syntaxe d'une requête sémantique dans SMW (a) et l'affichage de son résultat dans OncoLogiK (b).

cluses directement dans le wikitext d'une page, pour la création de listes dynamiques ou pour la recherche de valeurs. Elles interrogent les propriétés sémantiques définies dans les pages.

La figure 3.7 donne un exemple de requête dont le but est d'obtenir la liste des référentiels dont la date de mise à jour est supérieure à 2 ans. La liste est triée selon la valeur de la propriété `a pour date de mise à jour`, et affichée en fonction de modèles.

Dans le cadre d'OncoLogiK, les requêtes sémantiques sont utilisées de trois manières :

- Dans des pages de maintenance, pour créer des listes utiles telles que celle de l'exemple précédent ;
- Pour contrôler l'affichage de certaines catégories : par exemple la catégorie « Appareil digestif » contient les référentiels publics « Ampullome », « Anus », « Estomac », etc. Mais également les référentiels de la zone de travail « Travail:Ampullome », « Travail:Anus », « Travail:Estomac », etc. Dans ces cas, pour simplifier la navigation, des requêtes sémantiques sont utilisées pour n'afficher que les référentiels publics.
- Dans le cadre des *templates*, en particulier ceux intégrés à des fonctions en rapport avec les annotations MeSH. À partir d'un terme MeSH annotant un référentiel, des requêtes sont utilisées au besoin pour récupérer le code MeSH correspondant ou la traduction anglaise.

3.4.2 Interroger des ressources externes

3.4.2.1 Génération de requêtes vers les site de publication

Les GBPC édités sur OncoLogiK sont largement inspirés de la littérature médicale spécialisée. Ils peuvent être vus comme un état de l'art du domaine concerné au moment où ils sont réalisés. Toutefois, ils ne peuvent prétendre à l'exhaustivité, en particulier parce qu'ils contien-

nent des informations liées aux cas les plus génériques. Ainsi, les praticiens les consultant ont souvent besoin d'informations complémentaires. Une partie de celles-ci peut être présente dans les références bibliographiques fréquemment intégrées à la fin de chaque GBPC. Mais, dans de nombreux cas, la lecture d'autres sources documentaires est nécessaire, en particulier celles plus récentes que le GBPC.

PubMed et CISMeF, présentés dans la section 1.4.2.2, constituent des sources documentaires intéressantes dans ce cadre, indexant chacun de larges quantités de documents médicaux. Dans les deux cas, cette indexation est réalisée avec des termes du thésaurus MeSH, pris en compte par leur moteur de recherche interrogeable à distance. La recherche d'information dans ces sites peut toutefois s'avérer complexe de par la quantité de documents indexés, en particulier pour les personnes ne maîtrisant pas le thésaurus MeSH. Une premier orientation des utilisateurs peut être faite en sélectionnant les termes MeSH relatifs à un GBPC et en les transmettant au moteur de recherche. La réponse fournie constitue alors une bibliographie contextualisée relative au GBPC consulté.

Un tel mécanisme a été mis en place sur OncoLogiK, reposant sur les termes MeSH indexant les référentiels. D'un point de vue technique, il repose notamment sur l'emploi de requêtes sémantiques et de modèles. Son fonctionnement est présenté dans la figure 3.8. Chaque GBPC étant indexé avec des termes MeSH, des requêtes sémantiques sont créées dans le but d'obtenir la version du terme compréhensible par le moteur qu'on souhaite interroger. Celle-ci est présente dans la page du terme sous forme d'annotation. Le résultat est alors utilisé dans des modèles dont l'objectif est de formater un lien hypertexte vers la ressource documentaire. Ces liens contiennent les informations nécessaires à l'interrogation du moteur. Ils sont affichés dans un cadre spécial dans la page du GBPC. Dans le cas de PubMed, plusieurs liens sont générés afin d'utiliser un filtre spécial appelé «*Clinical Queries*» qui permet de préciser les requêtes selon des catégories prédéfinies («*Etiology*», «*Diagnosis*», «*Therapy*», «*Prognosis*» et «*Clinical Prediction Guides*») correspondant aux cas d'utilisation des professionnels de la santé.

Dans le cadre de cette application, si un GPBC est annoté par plusieurs termes MeSH, ceux-ci sont reliés par les connecteurs logiques OU ou ET. Toutefois, les moteurs de recherche permettent une expressivité bien plus étendue en intégrant également le connecteur NON. L'emploi de l'ensemble des connecteurs permettrait d'établir des stratégies de recherche plus complètes, comme montré par Montori *et al.* [2005].

3.4.2.2 Interrogation de bases distantes

L'objet de cette partie est de montrer de quelle manière des bases du Web de données peuvent être exploitées afin de proposer une plus-value en terme de contenu. Dans ce cadre du thésaurus de pharmacologie, une extension de SMW reposant sur des requêtes SPARQL a été créée. Son principe est d'explorer des bases distantes par l'intermédiaire de requêtes construites à partir du nom de produit pharmacologique présenté dans une page. Une fonction d'affichage gère ensuite un cartouche présentant les informations recueillies. Les bases ciblées de cette recherche sont DrugBank, spécialisée dans la description des molécules pharmacologiques, et DBpedia, une base de connaissances généraliste.

Dans la plupart des cas, de nombreuses informations peuvent être obtenues uniquement à partir du nom du produit pharmacologique. Malgré ces résultats encourageants, le module n'est pas exploité actuellement par OncoLogiK pour deux raisons :

- Le langage des données constitue la première raison. Les résultats obtenus sont en anglais, contrairement au wiki qui est francophone.
- La deuxième raison est bien plus importante et plus complexe à résoudre : les données

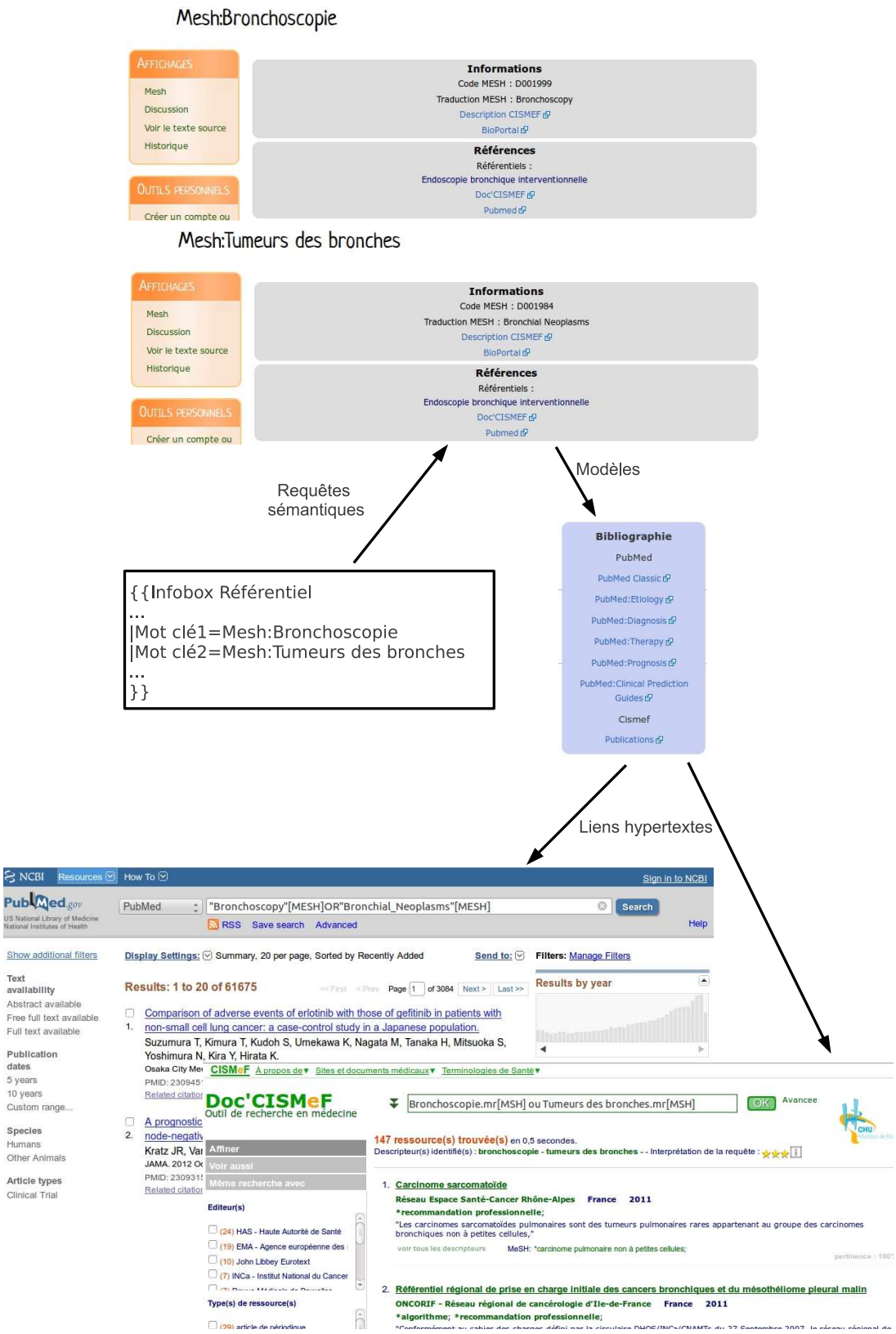


FIGURE 3.8 – Processus de création d'une bibliographie contextualisée.

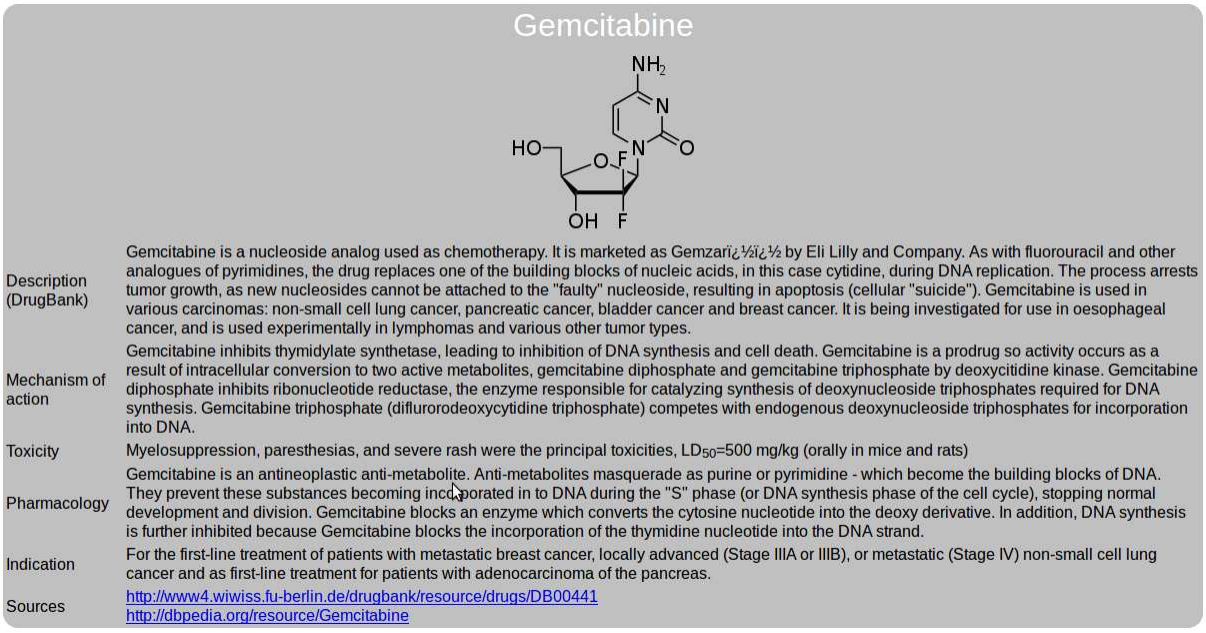


FIGURE 3.9 – Exemple de données importées depuis DBpedia et Drugbank concernant la gemcitabine.

présentées sont issues de bases distantes sur lesquelles les éditeurs d'OncoLogiK n'ont pas la main. Ce qui revient pour Oncolor à publier des données qu'ils ne peuvent pas contrôler et pour lesquelles aucune garantie de qualité n'est donnée. Cette démarche n'est pas compatible avec la charte HONcode à laquelle est soumise le wiki.

Pour pouvoir exploiter ce type de sources de données à l'avenir, des solutions sont envisageables. Pour contourner la barrière de la langue, une solution consisterait à utiliser des sources francophones. Malheureusement, on peut déplorer l'absence de sources de données en langue française. Toutefois, une version francophone de DBpedia [DBpedia.fr, 2012] a récemment vu le jour mais elle ne couvre actuellement qu'une partie des données de la version d'origine. En ce qui concerne la garantie de qualité des données, aucune base ne semble proposer de certification de leur contenu. Dans le cas de DBpedia, ce type de certification ne semble pas possible de par la nature même du projet fondé sur Wikipédia. Une solution alternative pourrait alors consister à importer ces données et à les certifier de manière interne avant de permettre leur affichage.

3.5 Évaluation par les utilisateurs

Pour permettre une première évaluation d'OncoLogiK, une enquête reposant sur des formulaires et des entretiens a été menée. Les personnes interrogées sont les quatre principaux contributeurs du wiki sémantique. Chaque personne a répondu Ã un questionnaire de manière individuelle, puis une réunion a été organisée pour partager et développer les réponses. Leurs rôles sont les suivnants :

- deux médecins en santé publique, responsables du suivi de l'édition des référentiels et de leur annotation,
- un infographiste, responsable de la mise en page,
- un coordinateur de rédaction des référentiels.

L'évaluation s'est déroulée autour de quatre thèmes. Le premier thème aborde les connaissances des personnes interrogées en matière de wiki avant le démarrage du projet, permettant ainsi d'en savoir plus sur leur niveau en matière de système collaboratif et leurs attentes. Durant la deuxième partie de l'évaluation, les questions ont concerné la formation et la prise en main de l'outil. Le troisième thème parle de l'appréciation du système, en particulier en comparaison avec l'ancien processus d'édition. Cette partie permet donc de déterminer les points forts et les points faibles d'OncoLogiK. Enfin, le dernier thème évoque les besoins des utilisateurs pour le futur.

Avant le démarrage du projet. Les contributeurs connaissaient la notion de wiki à travers Wikipédia, mais n'avaient aucune expérience d'édition de page sur de tels systèmes. S'ils étaient intéressés par l'idée d'utiliser un outil de travail collaboratif, certaines appréhensions existaient. En effet, l'intégration d'un système d'information collaboratif nécessite à leurs yeux un accompagnement pour les utilisateurs et des périodes de formation. De plus, si l'ancien système n'était pas optimal et semblait dépassé par certains aspects, il était opérationnel et fournissait des résultats enviés par d'autres réseaux régionaux. Ainsi, certaines fonctionnalités fournies par l'ancien système sont généralement absentes des moteurs de wiki telles que la création de versions imprimables de bonne qualité, de versions PDF et l'édition d'arbres de décision. La nécessité de disposer d'espaces de travail non publics a également été évoquée. Enfin, afin de justifier l'investissement dans un nouveau système, le wiki doit pouvoir mettre en avant de la valeur ajoutée par rapport à l'ancien système. Cette valeur ajoutée peut être amenée par l'ajout de fonctionnalités dont la réalisation n'aurait pas été possible avec un site Web 1.0.

Formation et prise en main de l'outil. Trois des contributeurs estiment à moins d'une demi-journée le temps nécessaire de formation pour pouvoir contribuer de manière efficace au wiki. La dernière personne évoque toutefois la difficulté de certaines balises spéciales (comme pour la gestion des références) ou complexe (par exemple, la construction de tableaux). Globalement, la navigation et l'édition de pages sont jugées assez simples. Les difficultés d'utilisation se situent surtout au niveau de fonctions plus complexes comme la gestion de l'historique, les interactions entre utilisateurs et l'administration du système. La période de transition entre les deux outils a donné lieu à un supplément de travail d'édition, puisque les modifications des GBPC devaient être faites en double. Grâce au soutien mis en place pour les éditeurs, cette période a servi de période de formation et a permis de comparer les deux modes d'édition et de constater les avantages et les inconvénients du wiki.

Utilisation au quotidien. Après son ouverture au public, un premier bilan peut être tiré de l'utilisation d'OncoLogiK. Les utilisateurs se déclarent globalement satisfaits de leur nouvel outil et listent plusieurs de ses avantages. Le premier élément de satisfaction est la standardisation permise par le wiki. Celle-ci uniformise la présentation des référentiels, ce qui améliore leur qualité. Elle permet également un gain de temps dans la mise à jour des GBPC. Les facilités d'édition permettent ainsi de moins se concentrer sur la forme du document et plus sur son contenu. Toutefois, un utilisateur souligne que cette standardisation a pour conséquence de permettre moins de liberté et de souplesse dans l'affichage.

Au niveau des capacités d'édition, la syntaxe wiki correspond aux attentes des utilisateurs pour la mise en forme des contenus. Cette syntaxe est jugée assez simple pour être apprise rapidement et utilisée par le public visé. La seule difficulté rencontrée est la création de tableaux. Un utilisateur souligne que leur création était plus simple en utilisant les éditeurs WYSIWYG

de la solution précédente.

De plus, la navigation dans l'ensemble du site et à l'intérieur des GBPC a été améliorée grâce aux fonctionnalités des wikis sémantiques. En effet, certaines fonctionnalités présentes de manière standard dans les wikis n'étaient pas présentes dans l'ancien site, comme le moteur de recherche ou la création automatique de sommaires. La catégorisation est également perçue comme un élément permettant d'améliorer la navigation et de clarifier la structure des informations. Enfin, les fonctionnalités sémantiques sont appréciées par les utilisateurs, qui voient en elles une plus-value apportée au système d'information. Toutefois, l'annotation des référentiels par les termes MeSH reste un travail complexe et coûteux en temps.

Au niveau du processus d'édition, la simplicité du système permet un gain de temps qui a pour conséquence d'améliorer la réactivité dans le processus de mise à jour. De plus, le travail de mise en page peut désormais être réalisé à distance, ce qui améliore le confort de travail. Enfin, le point le plus important dans ce contexte est que les utilisateurs soulignent que le système permet un réel travail collaboratif qui peut intégrer des experts médicaux aux compétences informatiques et aux disponibilités limitées.

Évolutions souhaitées. Lors de la conclusion de l'entretien, les utilisateurs ont émis des demandes pour la suite du projet. Certaines ont déjà pu être satisfaites, en particulier sur des réglages techniques du wiki ou sur l'ajout d'un éditeur d'arbres de décision, dont le fonctionnement est présenté dans le chapitre 5. D'autres restent encore à satisfaire. En particulier, le besoin de formations complémentaires sur les fonctions plus complexes du wiki est souligné, par exemple sur la gestion des utilisateurs et sur la manière d'utiliser l'historique des pages. D'autres propositions visent à simplifier l'édition. En particulier, les utilisateurs souhaiteraient disposer d'un éditeur WYSIWYG pour la création des tableaux, mais également d'un outil pour les guider dans l'annotation des référentiels par termes MeSH.

3.6 Conclusion partielle

Installé depuis juillet 2011, OncoLogiK est en production depuis avril 2012, après la migration, la correction et la validation des GBPC. La satisfaction des utilisateurs montre l'apport du wiki sémantique par rapport au système précédent. En particulier, les outils collaboratifs facilitent la gestion et la mise à jour des GBPC et les services sémantiques proposent une plus-value pour la navigation. À ce stade toutefois, les connaissances extraites des GBPC sont assez limitées : elles concernent uniquement les méta-informations (date, auteur, description, etc.). On peut regretter pour le moment que les connaissances décisionnelles, principal objet de ce travail, ne soient pas extraites. En observant les guides, on peut remarquer que beaucoup de ces connaissances sont décrites dans les GBPC sous forme d'arbres. Ces arbres sont intégrés au wiki dans des images générées par des logiciels commerciaux. Le niveau de collaboration sur le wiki pourrait être amélioré sur une application en ligne permettant d'éditer ces arbres de manière collaborative. D'un point de vue de la gestion des connaissances, avoir la main sur un tel éditeur permettrait d'extraire les connaissances décisionnelles contenues dans ces arbres.

Dans ce cadre, il est d'abord nécessaire de déterminer le langage employé afin de décrire ces connaissances décisionnelles. Dans la section 1.2.3 est évoquée l'approche fondée sur les connaissances pour les guides de bonnes pratiques informatisées. Dans le cadre de cette approche, plusieurs formats pour les connaissances décisionnelles en médecine ont été proposés. Le but du chapitre suivant est de présenter ces différents formats.

Chapitre 4

Connaissances décisionnelles en médecine

Sommaire

En informatique médicale, les connaissances décisionnelles sont souvent décrites par des GBPC. Ces documents contiennent un ensemble d'informations décrivant un processus de traitement médical : l'intention du GBPC, la description du contexte médical, les critères d'éligibilité du patient, des recommandations de traitement ainsi que les preuves sur lesquelles elles reposent, et des références bibliographiques. D'une certaine manière, ils peuvent être vus comme des « livres de recettes » pour les praticiens. Afin de pouvoir les exploiter automatiquement, une étape de formalisation est nécessaire, permettant de produire des GBPI. De nombreux formats sont apparus au cours des quinze dernières années pour les représenter. D'après Shiffman *et al.* [2000], ces formats doivent remplir six critères :

- être capables d'exprimer toutes les connaissances contenues dans un GBPC,
- proposer assez d'expressivité pour exprimer la complexité et les nuances de la médecine clinique,
- être flexibles, c'est-à-dire proposer différents niveaux de granularité de manière à ce que les recommandations puissent être interprétées à différents niveaux d'abstraction,
- être compréhensibles de manière à ce que les experts médicaux puissent les utiliser sans effort,
- pouvoir être partagés entre les institutions,
- être réutilisables dans toutes les phases d'utilisation du GBPI, c'est-à-dire lors de sa formalisation, de son exécution et de sa mise à jour.

L'objet de ce chapitre est de présenter les travaux effectués pour représenter les GBPI. La première section propose un panorama des formats existants en informatique médicale, tandis que la section suivante en propose une synthèse. La troisième section présente quelques propositions reposant sur les technologies du Web sémantique.

4.1 Des formats pour les GBPI

Plusieurs états de l'art sur les formats de GBPI sont disponibles dans la littérature. Ceux-ci forment les principales sources de la section suivante. Ainsi, les plus exhaustifs sont ceux de Georg [2006] et de Kaiser et Miksch [2005], présentant chacun un nombre conséquent de formats. En outre, le site OPEN Clinical [OpenClinical, 2013] fournit un survol complet des méthodes et outils disponibles pour chacun des formats. D'autres publications étudiées présentent uniquement les formats considérés comme les plus représentatifs par leurs auteurs. En particulier, Peleg *et al.* [2003] comparent six modèles (Asbru, EON, GLIF, GUIDE, PRODIGY et PRO*forma*) de manière à identifier les consensus et les points de désaccord. De la même manière, cinq formalismes (syntaxe Arden, Asbru, EON, GLIF et PRO*forma*) sont présentés et discutés dans deux études permettant de comprendre l'évolution des formats [de Clercq *et al.*, 2008, 2004]. Isern et Moreno [2008] proposent un point de vue original en se concentrant plus particulièrement sur les moteurs disponibles pour l'exécution des GBPI formalisés. Enfin, des études plus globales sur l'emploi des GBPI sont proposées par Sonnenberg et Hagerty [2006] et Latoszek-Berendsen *et al.* [2010], mettant l'accent sur l'utilisabilité de ces formats dans le contexte clinique et proposant de nouvelles perspectives d'évolution.

La suite de cette section propose un tour d'horizon des formats disponibles, où sont développés de façon plus importante les plus représentatifs, et faisant la synthèse des études précédentes.

4.1.1 Arden Syntax

Présentation. Mise en application dès 1992, la syntaxe Arden [Clayton *et al.*, 1989] a pour but de permettre le partage et la standardisation des connaissances médicales. C'est probablement

```
maintenance :
        title : Alerte sur le taux d'hématocrite ;;
        filename : low_hematocrit ;;
        version : 1.0 ;;
        institution : CPMC ;;
        author : George Hrispcsak, M.D. ;;
        specialist : ;;
        date : 1993-10-31 ;;
        validation : testing ;;
library :
        purpose : Avertir en cas d'apparition ou d'aggravation d'une anémie.
        explanation : À partir du résultat d'un hémogramme, l'hématocrite est vérifié en testant
        s'il est inférieur à 30 ou au moins à 5 points en dessous de la valeur précédente.
        keywords : anémie ; hématocrite ;;
knowledge :
        type : data-driven ;;
        data :          blood_count_storage := event ('Hemogramme') ;
                        hematocrit := read last ('Hematocrite') ;
                        previous_hct := read last ('Hematocrite')
                                where it occured before the time of hematocrit ;
        evoke : blood_count_storage ;
        logic :
                        /* Vérifie que l'hématocrite est un nombre valide*/
                        if hematocrit is not number then
                                conclude false ;
                        endif ;
                        if hematocrit <= previous_hct-5 or hematocrit<30 then
                                conclude true ;
                        endif ;;
        action :
                        write "L'hématocrite du patient est basse ou en chute" ;
```

FIGURE 4.1 – Exemple de MLM en syntaxe Arden, traduit de de Clercq *et al.* [2004].

le format le plus ancien et le plus répandu pour l'édition de GBPI. Contrairement aux autres formats présentés, Arden ne propose pas de de représentation graphique mais uniquement un pseudo langage de programmation assez simple, destiné à être compris par des experts médicaux. La version 2 [Sailors et Jenders, 2002] a été adoptée comme standard par les organismes de normalisation HL7 [HL7, 2013b] et ANSI [ANSI, 2013] en 1992.

Cette syntaxe, assez simple pour permettre une utilisation par des cliniciens, est utilisée pour décrire des modules logiques médicaux [39] (MLM). Un MLM est un module indépendant décrivant une prise de décision médicale. Les MLM sont composés de 3 *slots* :

- *Maintenance* contient des informations descriptives sur le module telles que le titre ou l'auteur,
- *Library* décrit l'objectif du MLM, explique son utilité et son fonctionnement et donne éventuellement des mot-clés le décrivant et des liens vers la littérature médicale,
- *Knowledge* est lui-même divisé en 4 *slots* : le premier (*Data*) indique les données nécessaires à l'exécution et correspond à la déclaration et à l'initialisation des variables utilisées, le deuxième (*Evoke*) renseigne le contexte de déclenchement qui active le MLM et qui peut être un événement ponctuel ou récurrent, le troisième (*Logic*) contient les tests logiques à effectuer pour déduire si la recommandation doit être faite et le quatrième (*Action*) décrit l'action devant se produire si les tests logiques du *slot* précédent retournent vrai.

Les tests logiques inclus dans les MLM sont des tests simples du type « *Si ... Alors ... Sinon ...* ». Seuls les types de données simples sont acceptés (*null*, booléen, entier, chaîne de caractères,

39. En anglais, *Medical Logic Modules*.

durée) : les types objet ne sont pas supportés. La figure 4.1 propose un exemple de MLM en syntaxe Arden.

Implémentation. Selon Sonnenberg et Hagerty [2006], la syntaxe Arden est considérée comme la plus établie dans le domaine de l'aide à la décision. Il existe plusieurs éditeurs de textes spécialisés dans l'aide à la création de modules Arden. Du point de vue de l'exécution des MLM, aucune spécification n'est donnée pour l'implémentation de moteurs. Il en existe plusieurs, dont la plupart sont intégrés à des solutions commerciales [OpenClinical, 2013].

Limites. Précurseur en matière de représentation des connaissances médicales, la syntaxe Arden est difficilement comparable avec les formats qui lui succèdent. En effet, elle ne permet pas la formalisation de GBPI entiers, mais est plutôt considérée comme un langage de création d'alertes.

De plus, l'absence de terminologie standardisée freine l'échange de données, car chaque structure hospitalière crée des MLM avec son propre vocabulaire. Plusieurs travaux visent à corriger ce manque, notamment Jenders *et al.* [2003] qui proposent d'utiliser le standard RIM [HL7, 2013a] du HL7 et Choi *et al.* [2006] qui essaient d'intégrer une ontologie médicale.

Une autre limite se situe au niveau des transitions entre MLM : si la syntaxe Arden prévoit des appels et des événements partagés, elle ne donne aucune restriction vis-à-vis de leurs implémentations. Chaque institut utilisant ce standard a donc défini son propre système. Ces implémentations disparates rendent les modules difficiles à partager entre instituts. De plus, ce problème complique l'écriture de modules issus de référentiels de grandes tailles, ou simplement préconisant plusieurs recommandations. Starren *et al.* [1994] montrent cette complexité dans le contexte de la gestion du suivi du patient post-opératoire après une chirurgie de l'artère coronaire.

4.1.2 Asbru

Asbru est un formalisme développé dans le cadre d'un projet pluridisciplinaire [The ASGAARD Project, 2013] regroupant des membres issus de plusieurs universités (Vienne, Stanford, Newcastle, etc.). Dans le cadre de ce projet, les GBPC sont vus comme des schémas génériques qui représentent les connaissances procédurales cliniques et qui sont instanciés et précisés dynamiquement par les praticiens de la santé sur des périodes de temps significatives. Pour formaliser les GBPI, le format Asbru exploite particulièrement les notions de squelette, de plan et d'intention.

Squelettes de plan et intentions. À l'origine conçu pour la représentation des protocoles médicaux, Asbru considère les GBPI comme une généralisation de ceux-ci. Shahar *et al.* [1998] parlent ainsi de squelettes de plan [40]. Ces squelettes sont des schémas de plans à différents niveaux de détails qui capturent l'essence d'une procédure. Créés dans le contexte de l'atteinte d'un objectif médical particulier, ils accordent une certaine flexibilité aux temps d'exécution de manière à être assez génériques pour être adaptés à différents contextes.

Chaque plan est conçu selon une certaine intention. La notion d'intention pour les GBPI est détaillée par Shahar *et al.* [1996]. Le but est de clarifier le GBPI en y indiquant les intentions explicites du rédacteur. Ces intentions peuvent être des actions à effectuer par les praticiens ou des objectifs à atteindre concernant l'état du patient. Un exemple d'objectif d'un plan pourrait être d'augmenter la concentration d'oxygène dans le sang.

40. «*Skeletal plans*»

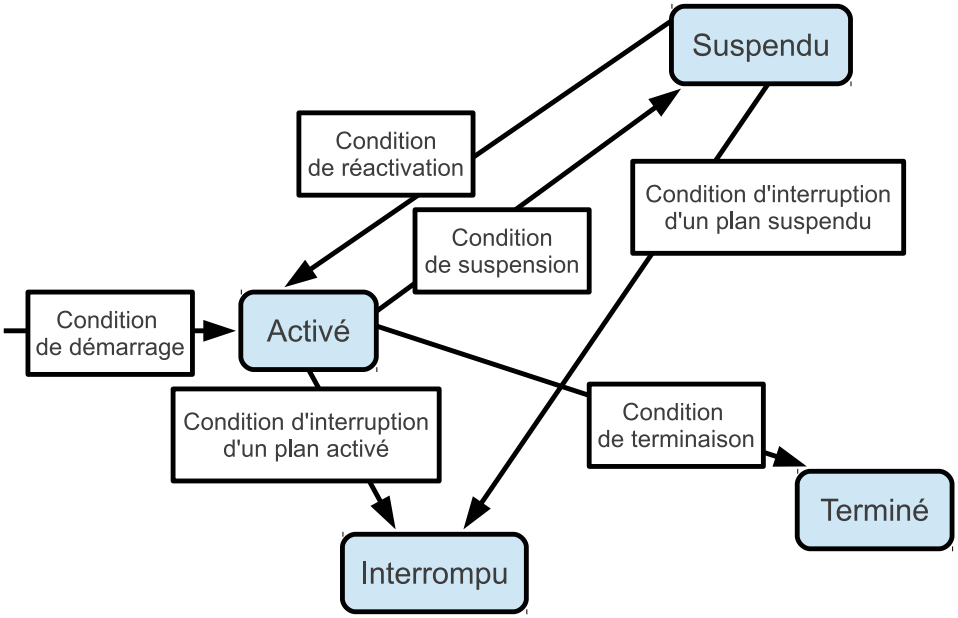


FIGURE 4.2 – Représentation des états des plans dans Asbru et de leurs conditions de transition [Kaiser et Miksch, 2005].

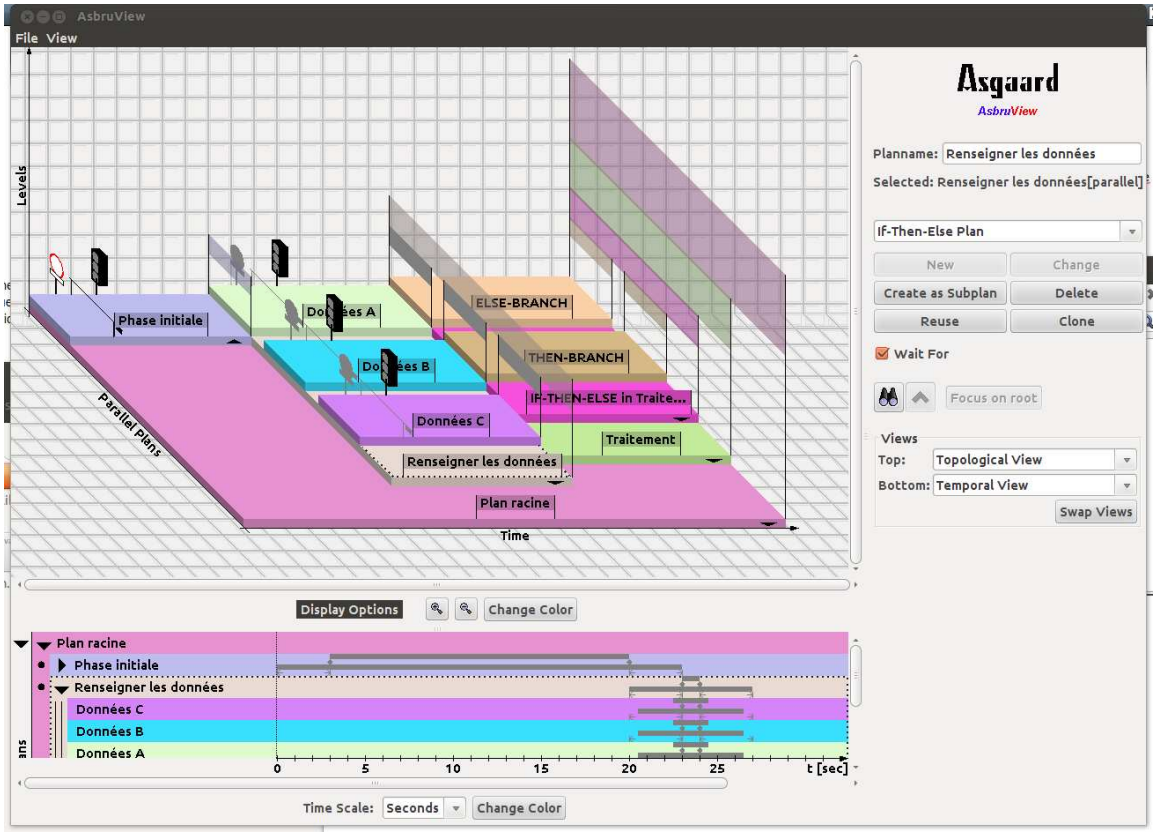


FIGURE 4.3 – Édition d'un GBPI avec AsbruView [Kosara et Miksch, 2001].

Contenu d'un plan. Dans Asbru, pour chaque plan, un certain nombre de « *knowledge roles* » doivent être définis :

- Les préférences sont les méthodes ou les contraintes à prendre en compte pour parvenir à réaliser l'intention. Un exemple de préférence peut être d'éviter une intervention chirurgicale.
- Les intentions sont les buts à atteindre tels qu'ils sont décrits dans le paragraphe précédent. Les intentions sont décrites par des modèles temporels pour le praticien. Quatre catégories d'intention sont définies :
  - les « *Intermediate states* » indiquent l'état du patient qui doit être maintenu, atteint ou évité,
  - les « *Intermediate actions* » détaillent les actes à effectuer lors de l'exécution du plan,
  - les « *Overall state patterns* » précisent l'état du patient qui doit être obtenu avant la fin du plan,
  - les « *Overall action patterns* » énumèrent tous les actes qui doivent avoir été effectués avant la fin du plan.
- Les conditions déterminent l'état courant du plan lors de l'exécution du GBPI. Comme le montre la figure 4.2, un plan peut prendre quatre états : activé, suspendu, interrompu ou terminé. Des conditions gèrent les transitions entre ces états.
- Les effets décrivent les relations entre les arguments du plan et les données médicales mesurables.
- Les structures de plan sont composées de plans et d'actions. Un plan peut ainsi être composé d'autres plans, qui peuvent être exécutés de manière séquentielle, parallèle, concurrente ou cyclique, et d'actions, qui correspondent à une unité de plan non décomposable.

Les aspects temporels constituent également une spécificité importante des plans Asbru. Asbru permet de définir des points temporels de référence dans l'exécution et de les utiliser pour la définition de durées.

Édition et exécution. Les GBPI en langage Asbru sont écrits selon une syntaxe formellement définie, proche de la syntaxe LISP [de Clercq *et al.*, 2004]. Cette syntaxe étant difficilement lisible pour les experts médicaux, leur visualisation peut également se faire via l'outil dédié AsbruView [Kosara et Miksch, 2001]. Contrairement à la plupart des autres formats, la visualisation n'utilise pas de graphe. La représentation repose sur des vues axées sur le déroulement temporel de l'exécution. La figure 4.3 montre un exemple de création de GBPI et les vues qui lui sont associées : la partie haute propose une « *Topological View* » tandis que la partie basse présente une « *Temporal View* ».

Le *framework* DeGel [Shahar *et al.*, 2003] propose également des logiciels permettant de publier et d'exécuter des GBPI au format Asbru. Son but est de proposer un outil facilitant la traduction des GBPC en GBPI.

4.1.3 EON/SAGE

Développé entre 1996 et 2003 à l'université de Stanford, le projet EON [Musen *et al.*, 1996] propose une architecture comprenant un ensemble de modèles et de logiciels pour la création d'applications ayant pour support les GBPI. Une partie des résultats a servi de base à son successeur, le projet SAGE [Tu *et al.*, 2004], démarré dès 2002.

La période EON. L'architecture d'EON propose en particulier un éditeur de GBPI (s'appuyant sur Protégé) et un moteur d'exécution. Le modèle de données défini dans le *framework*,

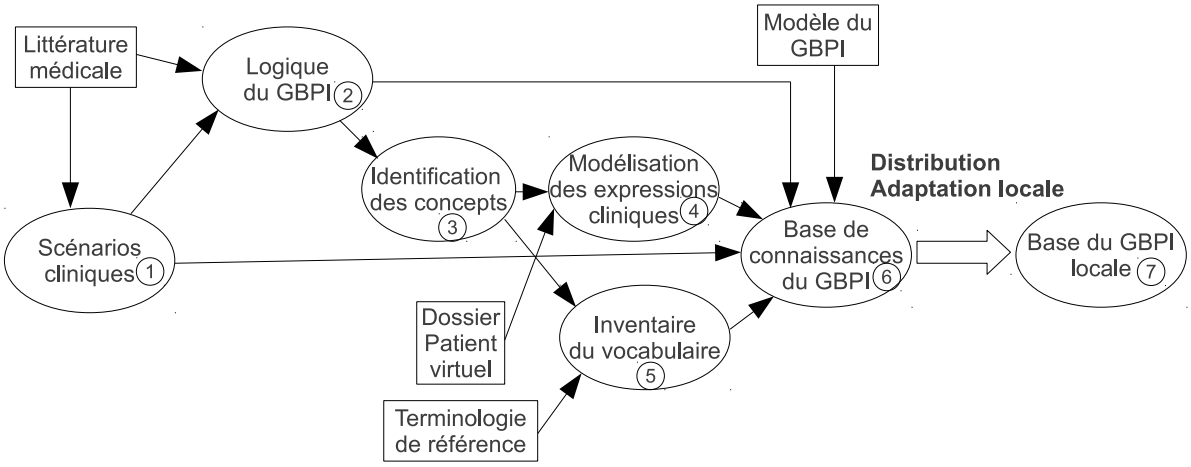


FIGURE 4.4 – Méthodologie de développement d'un GBPI avec SAGE (traduit de Tu *et al.* [2004]).

nommé Dharma, compte un ensemble de composants assez complet :

- un modèle temporel permettant de définir par exemple des durées, des points temporels et des intervalles,
- un modèle des concepts généraux de la médecine,
- un modèle de dossier patient simplifié,
- un langage d'expressions logiques,
- deux modèles de GBPI : le premier (*Consultation Guideline*) intègre des décisions et des actions indépendamment des considérations temporelles, à l'inverse du second (*Management Guideline*) qui prend en compte l'état et l'évolution du patient.

La présence de deux modèles amène de la flexibilité au *framework* en permettant une représentation à différents niveaux de granularité.

Un GBPI EON est représenté sous forme de graphe orienté contenant un ensemble de scénarios (ce qui correspond à des descriptions de l'état du patient), d'actions, de décisions, des contrôles de répétition et des mécanismes d'embranchement et de synchronisation. Pour chaque GBPI, une intention est formellement définie, en terme d'équation booléenne. Une intention peut également être définie à l'échelle d'une action ou d'une décision.

La période SAGE. Développé à partir de 2002 par un consortium réunissant des établissements médicaux, des entreprises et des équipes de recherche universitaire médicales et informatiques, le projet SAGE se fonde sur les conclusions du projet EON et intègre des idées d'autres formalismes pré-existants tels qu'Asbru, PRO*form*a, GLIF3, GEM, GUIDE et PRODIGY. Bénéficiant de ces expériences, trois objectifs ont guidé la création du format [Tu *et al.*, 2004] :

- intégrer les standards existant en informatique médicale afin d'optimiser l'interopérabilité du format,
- inclure des connaissances organisationnelles de manière à identifier les informations et les ressources du *workflow* lié à l'exécution du guide,
- synthétiser le travail de modélisation afin d'encoder toute la connaissance utile pour l'aide à la décision et maintenir le lien entre les sources d'information et l'utilisateur final.

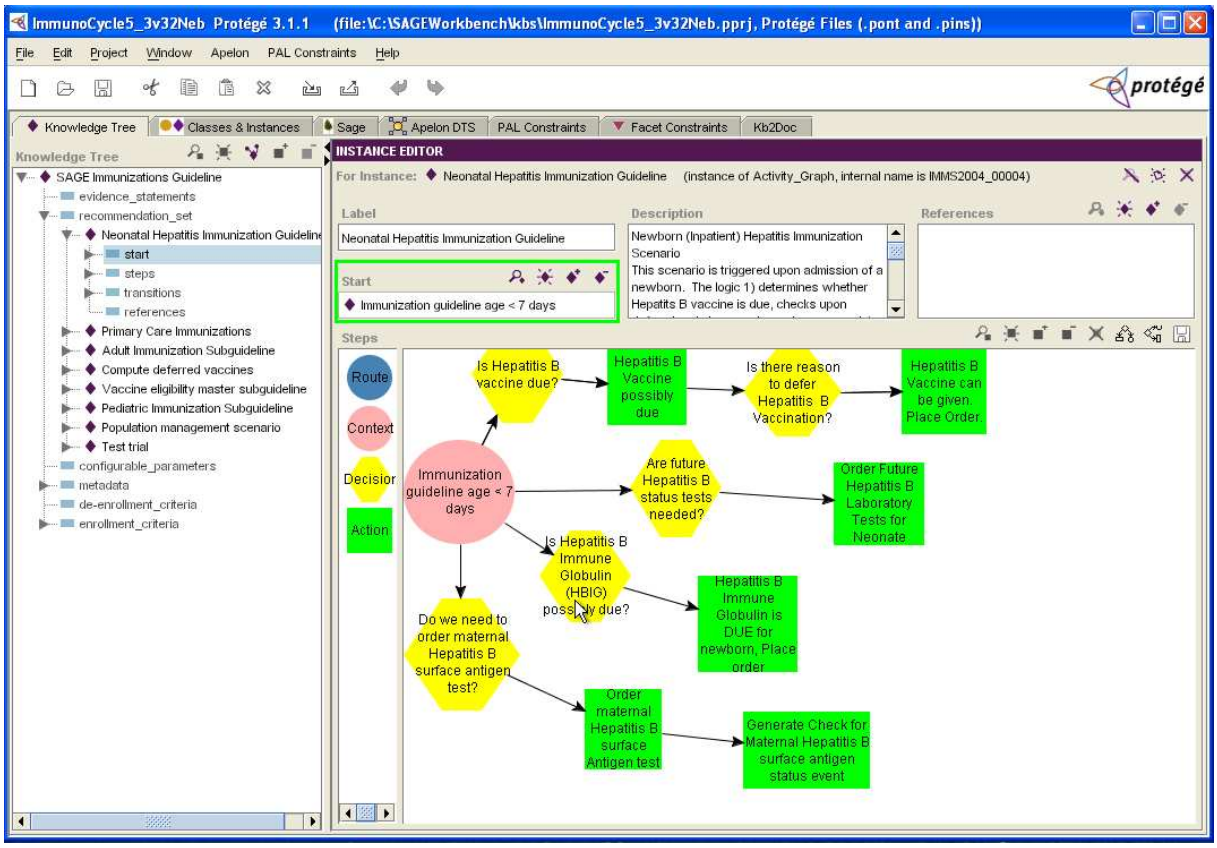


FIGURE 4.5 – Édition d'un GBPI sur l'immunité des nourrissons édité avec Protégé (Source : SAGE [2013]).

Ainsi, une méthode de développement des GBPI a été proposée, schématisée par la figure 4.4. Elle comporte sept étapes :

1. création de scénarios assez détaillés pour être représentés par un *workflow* ;
2. analyse des recommandations souhaitées et extraction de la logique et des connaissances nécessaires pour les générer ;
3. identification des concepts utiles ;
4. alignement des concepts avec le modèle de dossier patient ;
5. identification des concepts dans les ontologies du domaine (par exemple SNOMED CT) ;
6. traduction des scénarios et de la logique dans des modèles interprétables par des machines ;
7. adaptation de la base de connaissances au contexte local.

Pour créer le GBPI, le modèle propose des primitives classiques : *context* représente le contexte clinique et l'état du patient, *action* modélise une ou plusieurs activités du système de recommandation, *decision* sert dans le cas d'une acquisition de données ou d'une évaluation d'une transition, *routing* permet de synchroniser des chemins. Un GBPI peut être édité avec Protégé, pour lequel une extension a été proposée. Un exemple d'édition est présenté dans la figure 4.5.

Tu *et al.* [2007] présentent plusieurs des objectifs atteints par le format. La principale réussite est d'avoir montré qu'il est possible d'intégrer et de faire interagir plusieurs sources de données

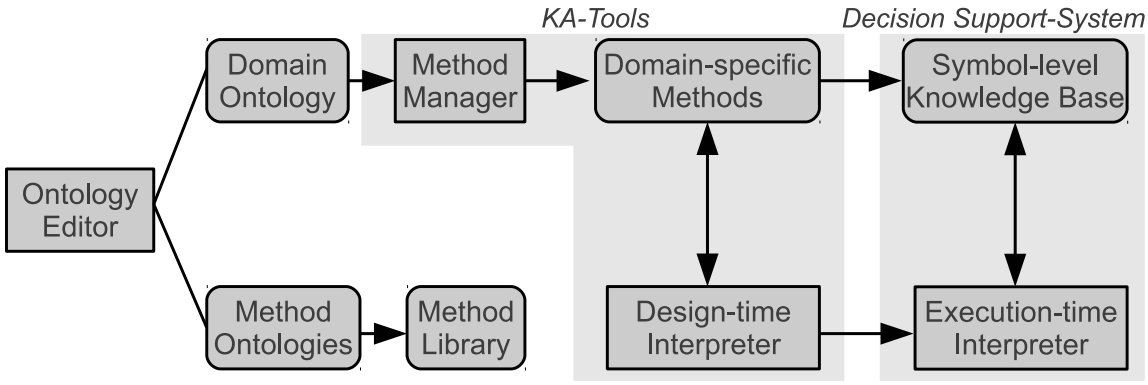


FIGURE 4.6 – Aperçu des composants du *framework* GASTON. Les rectangles correspondent à des logiciels et les arrondis à des modèles de données. Source : de Clercq *et al.* [2001].

(par exemple un dossier patient virtuel, une ontologie du domaine et un modèle de données des GBPI) au *workflow* représentant le GBPI. L'architecture proposée par SAGE permet de définir clairement le rôle de chacun. Les auteurs évoquent également la mise en œuvre du format : celle-ci reste circonscrite à des essais expérimentaux dans des institutions membres du consortium de développement.

4.1.4 GASTON

Développé au sein des universités de Maastricht et d'Eindhoven, GASTON [de Clercq *et al.*, 2001] est un *framework* constitué d'un ensemble de méthodes, d'outils et de modèles pour le développement des GBPI. Son but est de favoriser l'acceptation des GBPI et des systèmes d'aide à la décision dans les environnements en facilitant leur création.

La figure 4.6 présente une vue générale des outils et des modèles proposés par le *framework*. Ils sont utilisés dans une méthode en quatre étapes exposées par les auteurs :

1. Définir les ontologies de domaine et de méthode. Cette étape consiste à construire le vocabulaire qui sera utilisé par le GBPI. L'ontologie de domaine doit couvrir le domaine de spécialité du GBPI. Elle peut être créée spécialement ou être pré-existante, comme par exemple dans l'application détaillée par Dassen *et al.* [2003] où l'emploi de la CIM 10 est proposé. L'ontologie de méthode contient les primitives du langage de décision. L'emploi des primitives de GLIF2 (présenté dans la section suivante) est supporté par le *framework*.
2. Développer les librairies de méthode. Cette partie est destinée à recueillir les outils et méthodes de raisonnement automatique destinés à résoudre certaines tâches lors de l'exécution du GBPI. Les auteurs parlent ainsi de *Problem-Solving methods* pour qualifier un ensemble de méthodes d'intelligence artificielle telles que la classification, la satisfaction de contraintes, etc.
3. Éditer le GBPI. Le *framework* GASTON propose son propre outil d'édition, nommé KA-Tool. L'édition se fait sous forme graphique par la création de graphes orientés utilisant les primitives comme nœuds.
4. Exécuter le GBPI. Un moteur d'exécution a été spécialement développé dans le cadre de ce projet. Il a été optimisé de manière à proposer une aide à la décision en temps réel.

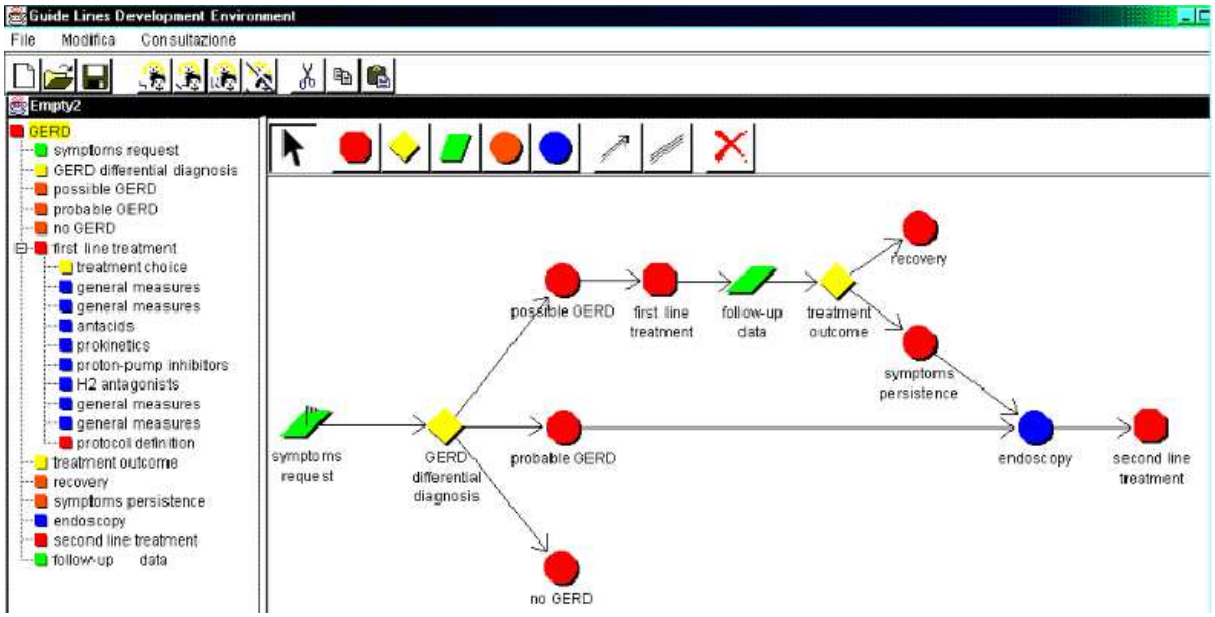


FIGURE 4.7 – Édition d'un GBPI au format GLARE avec *Guide Lines Environnement Developpement*, repris d' Anselma *et al.* [2007].

4.1.5 GLARE

Développé au sein du *Dipartimento di Informatica, Universiá del Piemonte Orientale «Amedeo Avogadro», Alessandria* (Italie), GLARE[41] [Terenziani *et al.*, 2003] propose un formalisme de représentation, un éditeur et un outil d'exécution des GBPI. Les GBPI au format GLARE reposent sur la notion d'action, distinguant *atomic actions* et *composite actions*. Les premières sont les briques de base du formalisme tandis que les secondes sont des éléments décomposables en d'autres actions. Ces actions sont liées entre elles dans des graphes orientés. Quatre types d'*atomic actions* sont définis :

- les *work actions* sont les actions devant être exécutées, décrites par un ensemble d'attributs (nom, étiquette, temps, ressources, buts),
- les *query actions* représentent les demandes d'informations,
- les *decisions* permettent de choisir un chemin spécifique dans le graphe,
- les *conclusions* représentent la sortie explicite du processus de décision.

Les *composite actions* sont définies par leurs composants à travers une relation de type « partie de ». Une autre relation établit la manière dont les composants sont exécutés. L'exécution peut être séquentielle, alternative, répétitive ou contrôlée par des contraintes temporelles.

Un éditeur de graphes GLARE nommé *Guide Lines Environnement Developpement* a été développé, dont l'interface est présentée à la figure 4.7.

Le traitement de contraintes temporelles dans GLARE permet de gérer la répétition et la périodicité des actions. Pour ce faire, plusieurs algorithmes à complexité polynomiale permettent de vérifier la consistance des contraintes temporelles [Stantic *et al.*, 2012].

41. *GuideLine Acquisition, Representation, and Execution.*

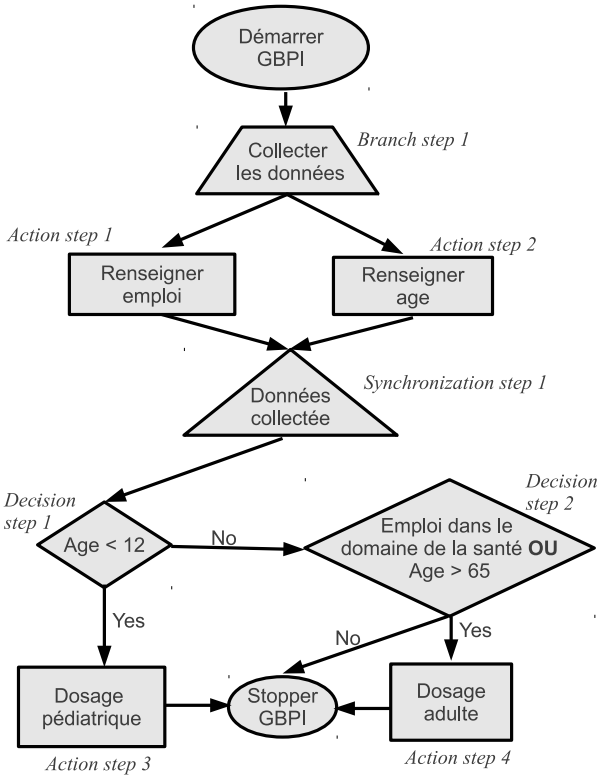


FIGURE 4.8 – Graphe dirigé au format GLIF sur le traitement de la toux, traduit de de Clercq *et al.* [2004].

4.1.6 GLIF

GLIF [42] est un format dont la première version publique, appelée GLIF2, fut publiée en 1998. Datée de 2004, sa version actuelle, GLIF3 [Peleg *et al.*, 2000], est présentée dans le reste de cette section. Son but principal est de proposer des GBPI conçus de manière à faciliter leur échange entre les institutions médicales en les modélisant sous une forme compréhensible par les experts humains et exploitable par les *parsers* utilisés dans les systèmes d'aide à la décision. Pour ce faire, GLIF propose de représenter les GBPC sous forme de graphes orientés consistant en un ensemble d'étapes ordonnées représentant des actions et des décisions. Développé jusqu'en 2003 par l'*Intermed Collaboratory*, le travail de recherche et d'implémentation a fait l'objet d'un groupe d'étude au sein du HL7 [43].

Modèle de données. Concrètement, un GBPI GLIF correspond à un ensemble d'étapes liées dans un graphe orienté. La définition du type de ces étapes est donnée par un modèle de données orienté objet défini en UML où sont décrites les primitives du langage (par exemple les décisions et les actions). Ce modèle définit 5 types d'étapes :

- Les *Decision steps* ont pour but de proposer des alternatives dans le déroulement du graphe. On peut en distinguer deux types :
  - les *Case steps* dont la sortie est définie selon une description logique,

42. *Guideline Interchange Format.*
43. *HL7 Clinical Guideline Special Interest Group.*

  - les *Choice steps* dont la sortie est déterminée en fonction du choix d'un agent externe, généralement un utilisateur.
- Les *Patient State steps* constituent des étiquettes pour décrire l'état courant du patient, souvent utilisées comme point d'entrée du graphe.
- Les *Branch steps* sont des arcs orientés multiples pointant vers des étapes à exécuter en parallèle. Elles permettent de définir des branchements vers plusieurs chemins différents.
- Les *Synchronisation steps* permettent de réunir des processus exécutés indépendamment, en synchronisant des chemins initiés par des *Branch steps*.
- Les *Action steps* représentent les actes médicaux proposés par le GBPI en fonction du cas décrit. Il en existe trois types :
  - les recommandations de traitement,
  - les actions utiles à la continuation de l'exécution du GBPI, comme par exemple la collecte de données,
  - l'appel à d'autres structures ou processus, comme l'exécution de macros ou de sous-parties du GBPI.

Un exemple de graphe utilisant ces étapes est donné par la figure 4.8.

Niveaux d'édition. Les standards de GLIF3 proposent trois niveaux d'implémentation pour chaque GBPI. Les niveaux sont conçus de manière à optimiser la lisibilité, l'exécution et l'adaptation des GBPI. Ces niveaux sont les suivants :

- *Conceptual Level* : ce niveau est prévu pour permettre l'édition par des experts médicaux et une meilleure lisibilité directe. Concrètement, les GBPI prennent la forme des graphes orientés présentés dans le paragraphe précédent.
- *Computable Level* : ce niveau utilise le modèle de données UML pour décrire les GBPI à un niveau formel. Ils prennent la forme de fichiers RDF/XML. L'implémentation est alors le travail d'un ingénieur des connaissances. Le modèle est prévu pour intégrer l'emploi d'un vocabulaire controlé développé par le HL7, RIM [HL7, 2013a].
- *Implementable Level* : le but de ce niveau est de prévoir l'interface avec les informations nécessaires du GBPI dans un contexte local, comme par exemple l'interfaçage avec le dossier patient.

Ainsi, les deux premiers sont ceux destinés à favoriser l'échange des GBPI entre les établissements médicaux, alors que le troisième prend en compte les particularismes locaux. Il est à noter que l'édition des trois niveaux est indépendante d'un point de vue technique. En effet, il n'existe pas de processus automatique permettant de développer un niveau à partir du précédent.

Outils d'édition et d'exécution. Les concepteurs de GLIF3 préconisent l'utilisation de l'éditeur générique de bases de connaissances Protégé [Stanford, 2013]. Ainsi, des extensions dédiées au format ont été proposées. Toutefois, elles correspondent à des versions dépassées de Protégé et ne sont plus disponibles en ligne.

Le moteur GLEE, spécialement conçu pour l'exécution des GBPI au format GLIF3, est présenté par Wang *et al.* [2004]. Dédié à l'exécution locale dans un contexte clinique, il propose notamment des outils d'interfaçage avec le dossier patient et d'autres outils standards des systèmes d'information médicale. On peut également noter que le moteur d'exécution DeGel, évoqué dans la partie concernant le langage Asbru, propose également une compatibilité avec le format GLIF3.

Applications. Des cas de mise en œuvre de GLIF3 ont été présentés [Choi *et al.*, 2007; Peleg *et al.*, 2006]. La première étude rapporte la création d'un GBPI à partir d'un GBPC existant sur le suivi de la dépression tandis que la seconde s'appuie sur un GBPC national américain sur les soins des pieds dans les cas de diabète. Alors que la première semble ne concerner que les deux premiers niveaux d'édition (« *Conceptual Level* » et « *Computable Level* »), la seconde étude montre la manière dont les trois niveaux d'édition peuvent être implémentés. Les deux études reposent sur l'utilisation de Protégé pour l'édition et de GLEE pour l'exécution. Ce dernier fait également office d'environnement de test pour le GBPI. Au final, les deux études évoquent le besoin d'outils complémentaires pour faciliter l'intégration de services externes tels que le dossier patient et les systèmes d'aide à la décision.

4.1.7 GUIDE

GUIDE [Quaglini *et al.*, 2000] est développé au sein de l'université de Pavia (Italie). Cette approche consiste à considérer les GBPI et les protocoles de soins comme des *workflows*, c'est-à-dire « la modélisation informatique ou l'automatisation, en tout ou en partie, d'un processus métier » [44]. L'une des modélisations possibles des *workflows* se fonde sur le modèle mathématique des réseaux de Petri qui permettent de représenter des processus concurrents. Ce modèle a donné lieu à une adaptation pour correspondre à la représentation des GBPI, nommée *Careflow*.

L'implémentation a donné lieu à l'élaboration d'une architecture logicielle à 3 niveaux :

- *Guideline Management System*, qui contient les GBPI,
- *Electronic Patient Record*, c'est-à-dire le dossier patient,
- *Careflow Management System*, qui fournit le support d'organisation du *workflow*.

Cette séparation permet d'optimiser chaque partie de l'architecture en utilisant des technologies dédiées et de permettre aux experts d'intervenir dans le module correspondant à leur domaine de compétence.

4.1.8 HELEN

Développé au sein du *Heidelberg University Medical Center* (Allemagne), le *framework* HELEN [Universitatsklinikums Heidelberg, 2013] propose des outils, des modèles et une infrastructure pour l'édition et l'exécution des GBPI.

Ses objectifs sont les suivants [Skonetzki *et al.*, 2004] :

- représenter les GBPC de différentes origines en conservant leurs structures et leurs formats de présentation,
- proposer un éditeur simple pour les GBPI prenant en compte les futures adaptations locales,
- offrir une accessibilité Web aux GBPI,
- générer des recommandations automatiquement,
- favoriser leur échange entre les institutions.

Pour ce faire, HELEN propose un architecture complète orientée Web, incluant notamment un moteur d'exécution, un navigateur spécifique proposant différentes vues d'un même guide pour différents périphériques de sortie et un outil d'édition fondé sur Protégé.

Les primitives du langage sont décrites dans une ontologie. Le *framework* propose trois niveaux d'acquisition des informations du GBPI :

- le niveau *Pragmatics* contient des méta-informations sur le guide (auteur, version, etc.),
- le niveau *Adaptation* documente les contraintes pour l'adaptation locale,

44. « *The computerised facilitation or automation of a business process, in whole or part* » [Hollingsworth, 1995].

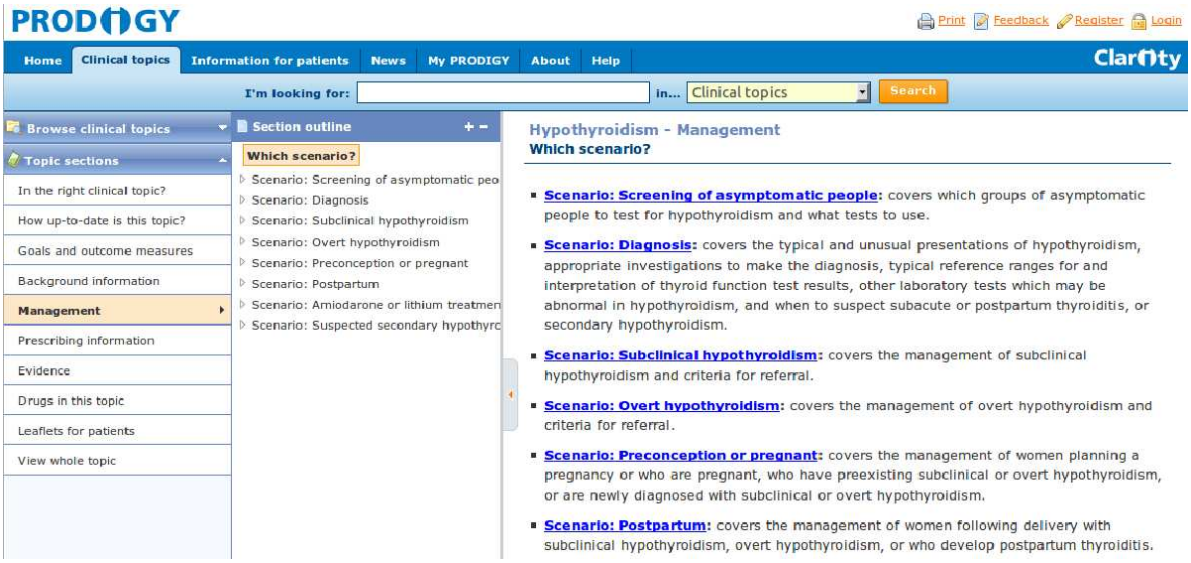


FIGURE 4.9 – Scénarios relatifs à la gestion de l'hypothyroïdisme dans PRODIGY [PRODIGY, 2011].

– le niveau *Document* contient les connaissances décisionnelles dans un ensemble de *Knowledge Modules*. Ces derniers contiennent en particulier la preuve sur laquelle ils reposent et une description formelle des connaissances, pouvant également être représentées sous forme de textes, d'images ou de graphes orientés.

Au stade de prototype, le *framework* ne propose ni évaluation, ni retour d'expérience en contexte clinique.

4.1.9 Prestige

Le projet Prestige [Gordon *et al.*, 1999] était un projet pour l'application de la télématique à la diffusion et l'utilisation des GBPI. La modélisation dérivait du projet DILEMMA [45], un modèle objet représentant les activités médicales et les agents impliqués. Son but était de définir un format électronique commun à tous les GBPI, indépendamment des spécialités médicales, des institutions et des applications logicielles.

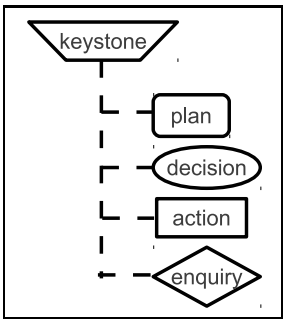
Le modèle proposé, le *Prestige Conceptual Guideline Model*, décrit les concepts et les relations nécessaire à l'exploitation de GBPI et de protocoles. Il est composé de deux parties, la première décrivant des généralités sur les soins médicaux et la seconde étant consacrée aux concepts liés à la création de protocoles. Le projet a ainsi permis de montrer qu'il était possible de créer un modèle de données portable pour les GBPI, c'est-à-dire un modèle applicable pour toute organisation, indépendant des technologies.

4.1.10 PRODIGY

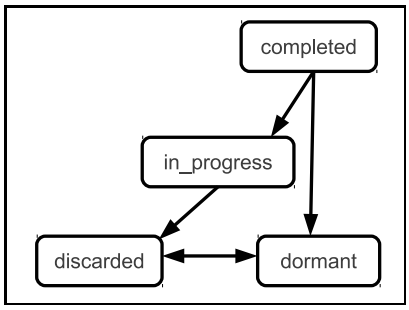
PRODIGY [46] [Johnson *et al.*, 2001] est un projet soutenu par le National Health Service (Royaume-Uni), créé pour le développement d'un système d'aide à la décision exploitant les

45. DILEMMA était un projet soutenu par l'Union Européenne pour le développement de l'aide à la décision, de 1991 à 1994.

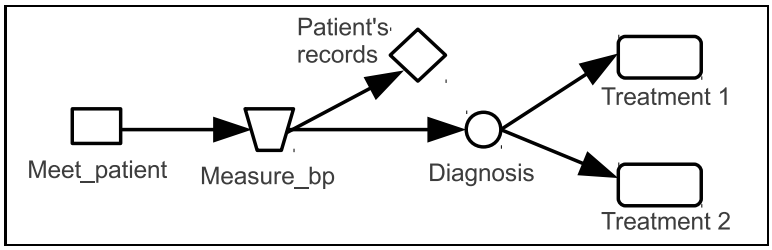
46. *Prescribe RatiOnally with Decision-support In General Practice studY*



(a) Hiérarchie issue de l'ontologie des tâches.



(b) Transition entre états des tâches.



(c) Exemple de GBPI au format PRO*forma* .

FIGURE 4.10 – Représentation des tâches dans PRO*forma* : hiérarchie de l'ontologie « PRO*forma* Task Ontology » (a), graphe de transition des états de ces tâches (b) et exemple de GBPI (c). Adaptés de Kaiser et Miksch [2005] et de Grando *et al.* [2012].

GBPC à destination des médecins généralistes. Dans PRODIGY, les GBPI sont modélisés comme un ensemble de scénarios, organisant la gestion des décisions et des actions. Les scénarios sont définis selon l'état du patient et son traitement. Pour chaque scénario sont définis un modèle de consultation décrivant textuellement les bonnes pratiques liées au scénario et un choix entre les différentes alternatives d'action à effectuer par le praticien. La figure 4.9 montre l'exemple de l'ensemble des scénarios décrits pour les cas d'hypothyroïdisme. Au final, la trajectoire d'un patient peut être vue comme un ensemble de scénarios successifs.

Le modèle de décision dans PRODIGY repose sur un ensemble de conditions d'entrée et de sortie des scénarios permettant d'associer au scénario courant la meilleure alternative disponible.

Les GBPI créés pour PRODIGY sont librement consultables en ligne et toujours maintenus à jour [PRODIGY, 2011].

4.1.11 PRO*forma*

PRO*forma* [Fox *et al.*, 1997] vise à proposer une approche pour soutenir la diffusion des GBPI sous une forme compatible avec les systèmes d'aide à la décision. Pour ce faire, l'approche propose un mécanisme d'aide à la décision et une gestion avancée des *workflows*.

Formalisme. Les GBPI PRO*forma* peuvent être représentés sous forme de graphes orientés. Les nœuds constituent des tâches dont la signification est décrite dans l'ontologie *PRO*forma *Task Ontology*, dont la hiérarchie de concepts est donnée dans la figure 4.10b.

Tous les concepts de tâches héritent de l'objet abstrait *keystone* qui n'a pas d'implémentation spécifique. Son but est de définir les attributs communs à ses quatre sous-concepts qui sont utilisés

lors de la création d'un GBPI :

- les *Actions* représentent les actions à effectuer dans l'environnement externe au GBPI (par exemple administrer un médicament ou mettre à jour une base de données),
- les *Enquiries* permettent de demander des informations à l'utilisateur ou à une application tierce,
- les *Decisions* correspondent aux choix thérapeutiques,
- les plans, qui constituent les briques de base du langage, correspondent à des ensembles de tâches.

Un exemple de GBPI sous forme de graphe dirigé est proposé dans la figure 4.10c. Le graphe utilise la signification des formes géométriques présentées dans la hiérarchie de concepts de la figure 4.10a. Le but est de guider un généraliste en lui proposant d'abord une action, c'est-à-dire rencontrer le patient. L'étape suivante propose de mesurer la pression sanguine. Elle est représentée par un *keystone* pour montrer que l'exécution peut être faite par le généraliste lui-même (dans ce cas, elle serait représentée par une figure de type *Action*) ou peut être déléguée (alors, elle serait représentée par une figure de type *Enquiry*). Ce choix sera déterminé lors de l'exécution du GBPI. Ensuite, deux étapes simultanées sont proposées : la première consiste à inscrire la valeur de la pression sanguine dans le dossier du patient tandis que la seconde propose au généraliste un diagnostic. En fonction de ce dernier, deux plans de traitement sont proposés.

Tous les types de tâche, qui héritent de la classe *keystone*, partagent des attributs. Chaque attribut correspond à une expression booléenne permettant d'obtenir une valeur de vérité utilisée dans l'exécution du GBPI. Les principaux attributs sont les suivants :

- l'attribut *State* permet de définir l'état d'exécution de la tâche, parmi les états présentés dans la figure 4.10b.
- *Precondition* et *Postcondition* définissent respectivement les conditions de début et de fin d'exécution et interviennent sur les processus de changement d'état décrits précédemment.
- *Antecedents_ task* précise l'ensemble des tâches qui doivent avoir été effectuées avant l'exécution de la tâche courante. Cet attribut peut être vu comme un cas particulier de *Precondition*.
- *Trigger* permet d'activer une tâche, quel que soit l'état de *Precondition*.
- *Goal* représente des conditions logiques qui, une fois vérifiées, entraînent l'arrêt de la tâche.

D'autres attributs sont définis afin de permettre l'exécution de cycle (*Nbr_ cycles_ until*, *Cycle_ until*), un besoin de confirmation de la part de l'utilisateur (*Confirmatory*), etc. De plus, chaque type de tâche dispose d'attributs qui lui sont spécifiques. Par exemple, les *Plans* ont un attribut *abort_ condition* qui permet d'annuler l'exécution des tâches qu'ils contiennent.

Exécution des GBPI. Dans PRO*forma*, l'exécution des GBPI se fait grâce à des moteurs dédiés. Ces moteurs ont en charge l'exécution de fonctions prédéfinies. Leur rôle est en particulier de calculer les valeurs des attributs et, selon les résultats, de déterminer l'état de chaque tâche. Deux environnements propriétaires pour le développement et l'exécution existent actuellement : Arezzo [Infermed, 2013] et Tallis [Cossac, 2013].

4.2 Synthèse des formats de GBPI

Dans la section précédente, de nombreux formats ont été présentés. Chacun propose des solutions à différents problèmes rencontrés lors de la formalisation des GBPI. Le but de cette section est de proposer une synthèse des technologies et des modélisations utilisées et de mettre en avant plusieurs points encore à résoudre.

Tableau récapitulatif. La table 4.1 résume les choix de leurs auteurs respectifs. La première colonne décrit les primitives de langage de représentation utilisées. Les primitives peuvent être vues comme les briques de base du langage. La colonne suivante présente comment les contraintes de transition entre les différentes primitives sont décrites graphiquement. Typiquement, des graphes sont utilisés, afin de fournir une vision claire et compréhensible aux praticiens. Ensuite, pour chaque format, sont repris les outils d'édition compatibles. Puis il est indiqué si ce format a été prévu pour être représenté par un *workflow*. L'avant-dernière colonne présente la manière dont est prévu l'interfaçage avec le dossier patient. Enfin, la dernière colonne présente les ressources terminologiques et ontologiques médicales prises en compte lors de l'édition du GBPI.

Modèles de données. L'architecture de donnée la plus répandue consiste à proposer trois modèles de données complémentaires :

- Le modèle de décision comporte les outils nécessaires à la modélisation du processus de décision, et peut être indépendant du domaine. Le paragraphe suivant décrit ces modèles.
- Le modèle du domaine comporte les informations inhérentes au domaine médical. Il peut être spécifique au format d'édition (comme dans EON et GLARE) ou reposer sur l'emploi de RTO (comme SNOMED et LOINC dans SAGE).
- Le modèle de dossier patient consiste à représenter les données liées à un patient qui seront utilisées lors de l'exécution du GBPI. Étant donné l'absence de standard, une grande variété de modèles existe pour représenter le dossier patient. Les formats proposent généralement une représentation simpliste correspondant à leur approche, facilitant les interactions avec les autres modèles utilisés.

La séparation la plus explicite est celle proposée dans le projet SAGE. L'une de ses principales réussites est d'avoir respecté ce découpage, en particulier en précisant leur utilisation à différents niveaux dans la méthodologie de création des GBPI. Dans leur conclusion du projet, Tu *et al.* [2007] évoquent d'ailleurs un goulot d'étranglement de l'acquisition des connaissances. Il est en effet difficile de maintenir et développer des bases de connaissances cohérentes pour les GBPI alors que le domaine manque de standard compréhensible. Lors de l'édition, l'utilisation de standards nécessite beaucoup de travail et un échange intensif entre les cliniciens et les ingénieurs des connaissances. Dans de nombreux cas, l'adjonction de connaissances additionnelles est nécessaire pour désambiguïser la représentation.

Modèles de décision. Globalement, on peut remarquer certaines similarités sur l'emploi des primitives du langage. Cette remarque est précisée par Wang *et al.* [2002] qui affirment que tous les modèles de représentation des GBPI doivent contenir des primitives de type action (c'est-à-dire une action interne ou externe au SSAD) et de type décision (c'est-à-dire un choix laissé à l'utilisateur). Les auteurs de SAGE exploitent cette proposition en étendant leur format à seulement deux autres primitives. Ces derniers parlent d'ailleurs de « léger consensus » sur leur sémantique et appellent à la reconnaissance d'un standard [Tu *et al.*, 2007]. Ainsi, chaque format utilise un ensemble fini de primitives, hormis EON qui propose un modèle extensible.

Seule la notion d'intention semble faire débat. Elle propose de déterminer comme un attribut essentiel du GBPI le but que l'on souhaite atteindre lors de son exécution. Lorsqu'elle est utilisée, elle est présente soit sous forme de texte (GLIF3), soit présentée de manière formelle par une équation booléenne (EON, Asbru, PRO*forma*).

Afin de représenter les contraintes de transition entre les états, chaque format, hormis Arden et Asbru, propose un modèle graphique reposant sur des graphes orientés. Cette représentation

Format	Primitives de représentation	Représentation graphique	Outils d'édition	Modélisation par *workflow*	Modèle de dossier patient	RTO exploitées
Arden Syntax	logic slots	Aucune	Outils commerciaux	Non	Non	RIM
Asbru	plan ; condition ; preference	Corps du plan	AsbruView	Non	Non	Aucune
EON	decision ; action ; scenario	Graphe orienté	Protégé avec extension	Non	Modèle Dharma	Modèle Dharma
GASTON	decision ; action	Diagramme de contraintes	KA-Tool	Oui	BD et système d'information	SNOMED
GLARE	decision ; action ; conclusion	Graphe de contraintes	Éditeur GLARE	Non	BD	Ontologie propre
GLIF	decision ; action ; patient state	Graphe orienté	Protégé avec extension	Non	Ontologie du dossier patient	RIM
GUIDE	decision ; task	Réseaux de Pétri	Guideline Authering	Oui	Modèle relationnel	SNOMED
HELEN	decision ; action	Graphe orienté	Protégé avec extension	Non	Modèle dossier patient	Aucune
Prestige	plan ; condition ; preference	Protocole	GAUDI, GLGAM	Non	Modèle dossier patient	GRAIL
PRODIGY	decision ; action ; scenario	Diagramme de transition	Protégé avec extension	Non	Ontologie du dossier patient	Non indiqué
PRO*forma*	decision ; action ; enquiry ; task	Graphe de contraintes	Arezzo, Tallis	Oui	Non	Non indiqué
SAGE	decision ; action ; contexte ; routing	Graphe orienté	Protégé avec extension	Oui	Ontologie du dossier patient	SNOMED CT, LOINC

TABLE 4.1 – Récapitulatif des formats de décision des GBPI.

est facilement compréhensible et ne nécessite aucune formation préalable pour sa compréhension. Leur large utilisation a créé des habitudes d'utilisation pour les médecins qui la considèrent désormais comme un standard. Pour sa part, Asbru propose de représenter différemment les GBPI en mettant en avant le déroulement du GBPI, intégrant de manière plus explicite les contraintes temporelles. Enfin, Arden ne propose aucune représentation graphique, mais des suites de modules indépendants difficilement lisibles. Si ce type de langage peut parfaitement fonctionner pour des représentations simples (du type alerte), il est difficilement utilisable pour la création de GBPI plus longs ou plus complexes.

Dans chaque format, un langage spécifique est proposé pour définir les critères de transition entre les états. Les transitions sont généralement exprimées sous forme d'équations booléennes. Quelques particularités sont toutefois intéressantes à noter :

- Le langage le plus répandu est GEL et son extension GELLO, utilisables dans le cadre des formats Arden et GLIF, proposés comme un standard par HL7.
- Certains langages intègrent des contraintes temporelles : c'est notamment le cas d'Asbru et PRO*form*a.
- Enfin, certains proposent une gestion de l'incertain. Ainsi, GEL supporte trois réponses pour une proposition booléenne (*true*, *false* et *unknow*). Le but recherché est de pouvoir continuer l'exécution d'un GBPI en cas de données manquantes.

Outils d'édition. Pour l'édition des GBPI, plusieurs approches proposent l'utilisation de l'éditeur de connaissances Protégé. C'est notamment le cas des projets HELEN et SAGE pour lesquels des extensions spécifiques au format ont été développées. Cette solution présente l'avantage de reposer sur une solution stable permettant de prendre en compte plusieurs modèles de données. L'intégration d'interfaces de dessin dans l'éditeur permet notamment de dessiner des graphes orientés, ce qui convient le mieux aux experts.

De par leurs spécificités, d'autres solutions sont contraintes de développer leurs propres outils. Ainsi, Asbru propose un outil d'édition permettant d'éditer un GBPI en fonction des contraintes temporelles. D'une autre manière, GUIDE propose un éditeur permettant de faciliter la construction du GBPI dans le contexte de représentation des réseaux de Pétri.

Enfin, plusieurs formats ne proposent que des solutions commerciales, dont l'évaluation est difficile (Arden Syntax, GASTON avec KA-TOOL, PRO*form*a avec Tallis et Arezzo). Pour Asbru, une licence propriétaire protège le format de sortie du GBPI, ce qui constitue un frein à sa portabilité.

Exécution. Comme le soulignent Isern et Moreno [2008], chaque format dispose de son propre moteur d'exécution. L'expressivité et les formats des modèles diffèrent, ce qui a conduit aux développements de moteurs sur mesure, de manière à optimiser les performances. Plusieurs moteurs (par exemple SAGE ou GASTON) permettent des communications avec des dossiers patient virtuels, ce qui permet d'intégrer les données du patient directement aux GBPI et facilite ainsi l'utilisation finale.

Pour certains formats, l'exécution se fait dans le cadre de *workflows* (par exemple GUIDE et SAGE). La force de ce formalisme est son habileté à supporter la modélisation de processus complexes concurrents, qu'ils soient séquentiels, parallèles ou itératifs. Prenant en compte l'ensemble des ressources disponibles, ils optimisent l'opportunité d'une bonne intégration des guides dans un contexte clinique.

Utilisation. Peu de projets évoquent des implémentations en dehors de leur contexte de développement. Il semble donc que l'affirmation de Sonnenberg et Hagerty [2006] selon laquelle « Il n'existe pour les GBPI ni *framework* dominant, ni système répandu qui soit en utilisation clinique en dehors de l'institution dans laquelle il a été développé » soit toujours d'actualité. Certaines exceptions peuvent toutefois être évoquées : GLIF3 propose plusieurs implémentations (notamment Peleg *et al.* [2006] et Choi *et al.* [2007]), Arden Syntax est considéré par de Clercq *et al.* [2004] comme un format répandu et le succès du site Web de PRODIGY en fait une référence pour les médecins généralistes britanniques.

Cette faible popularité est en partie liée au manque de portabilité des GBPI. Ceux-ci semblent difficiles à utiliser en dehors de l'institution dans laquelle ils ont été conçus. Plusieurs problèmes peuvent l'expliquer :

- La présence de nombreux formats non compatibles due à l'absence de standard ne facilite pas l'intégration aux systèmes d'information existants.
- Comme le montrent Kaiser et Miksch [2005], chaque approche se concentre sur un aspect de la formalisation, se référant à différents buts, différents publics et différentes applications.
- Les formats manquent de méthodes et d'exemples d'application pour l'adaptation locale, même si elle est généralement prise en compte de manière théorique.

4.3 Web sémantique et GBPI

La diversité des propositions et l'absence de format dominant a conduit à une absence de standard pour la formalisation des GBPI. Une des conséquences de cette absence est la faiblesse des échanges de GBPI entre les institutions, ce qui constitue pourtant un des enjeux majeurs de ceux-ci. L'intégration de formats établis et reconnus, tels que ceux du Web sémantique, peuvent au moins partiellement pallier ce manque. Après avoir montré l'intérêt d'utiliser les technologies du Web sémantique pour le GBPI, quelques exemples d'applications sont présentés.

4.3.1 Intérêt du Web sémantique pour les GBPI

Dans leurs conclusions, Latoszek-Berendsen *et al.* [2010] insistent sur l'apport que constitue l'emploi des méthodes formelles dans le cadre des GBPI. En particulier, trois avantages sont identifiés :

- Le guide peut être considéré comme un système en développement pour lequel des vérifications sont nécessaires. Par exemple, on peut vérifier s'il comporte des incohérences ou s'il remplit certains critères de qualité.
- S'ils sont formellement validés, les GBPI nationaux pourraient être considérés comme des standards infaillibles lors de leur adaptation locale.
- Les vérifications pourraient également être utilisées pour détecter les comportements non désirés et pour analyser pourquoi un processus réel ne correspond pas au contenu du guide.

Ainsi, dans le cadre de création de systèmes d'aide à la décision, plusieurs approches ont proposé des traductions de GBPC utilisant des méthodes formelles pour le raisonnement, comme le font les formats présentés dans les sections précédentes. On peut également citer le système ASTI [Lamy *et al.*, 2010]. Les CPG y sont traduits sous forme de règles du type « *if* conditions *then* prescribe ... » modélisées manuellement. Toutefois, ce système n'implémente pas de standard de représentation des connaissances, ce qui nuit à sa portabilité.

L'interopérabilité des systèmes et leur qualité peuvent être améliorés par l'utilisation de technologies sémantiques. Pour Doucette *et al.* [2012], plusieurs avantages à utiliser ces technologies dans le cadre de la formalisation des GBPI sont évoqués :

- La formalisation des connaissances permet de partager des données avec des systèmes d'information médicaux hétérogènes.
- Des règles d'inférences peuvent être définies sans recours à des enregistrements médicaux incomplets ou erronés, et être utilisées pour plusieurs scénarios.
- En utilisant un moteur d'inférences, la réponse à une requête contient une véritable preuve construite par le moteur, expliquant comment le système est arrivé à cette réponse.
- L'efficacité des approches sémantiques repose sur des modèles indépendants de la quantité de données disponibles, par opposition aux algorithmes d'apprentissage.

Afin de soutenir la diffusion des GBPI de qualité entre les institutions, l'utilisation de standards établis peut être un premier pas. Le Web sémantique propose des standards pour la représentation de connaissances tels que RDF et OWL. Casteleiro et Diz [2008] montrent la possibilité de les employer pour la formalisation des GBPC et détaillent un ensemble d'avantages à les utiliser. En particulier, l'intégration des services Web permet d'utiliser plusieurs sources d'information et de les exploiter par l'intermédiaire de moteurs d'inférences. L'efficacité de ces derniers est également montrée par Peleg *et al.* [2012] où les mécanismes de raisonnement utilisent des représentations de primitives classiques de GBPI telles que les actions et l'état du patient. Jafarpour *et al.* [2011] montrent comment le raisonnement avec OWL 2 peut servir à gérer des diagrammes de transition d'état à la manière de ceux utilisés par Asbru et PRO*form*a, en prenant notamment en charge les calculs de pré-condition et de post-condition.

4.3.2 Quelques exemples d'applications

La première proposition de formalisation de GBPC en OWL est développée par Kashyap *et al.* [2005]. Le processus de décision est modélisé par un ensemble d'étapes de type *definition* (où sont décrites l'état du patient) et *decision*. En complément, une base est définie, contenant un ensemble de règles du type « *Si... Sinon,...* » représentant les recommandations de prescription. Des exemples inspirés d'une application sur la gestion des lipides dans les cas de diabète sont donnés et une architecture d'exécution reposant sur un moteur d'inférence *ad-hoc* est proposé. Toutefois, aucune implémentation concrète n'est présentée.

Reposant également sur une modélisation par étape, Abidi et Shayegani [2009] proposent un modèle d'ontologie plus général, intégrant par exemple les méta-données (titre, auteurs, etc.). Dans ce cas, le modèle est principalement utilisé pour réaliser la synthèse de plusieurs GBPC sur un même sujet.

Des règles « *Si... Sinon,...* » sont également utilisées par Abidi [2007]. Un modèle général de représentation des GBPC est proposé, utilisant quatre primitives : *Patient_ type*, *Symptom*, *Risk_ factor* et *Recommendation*. Ces concepts sont liés par des propriétés spécifiques, comme par exemple *has_ symptom* qui relie *Patient_ type* à *Symptom*. Une application à la formalisation d'un GBPC sur le cancer du sein est proposée.

Un *framework* pour l'édition et l'exécution des GBPI est présenté par Hussain *et al.* [2007]. À la manière des méthodes documentaires exposées dans le chapitre 1, il cherche à ajouter une couche sémantique à des GBPC existants. Ce *framework* repose sur l'emploi des technologies du Web sémantique, définissant le contenu de trois ontologies :

- La première ontologie représente la structure du GBPC, utilisant le modèle de GEM (présenté dans le chapitre) ;
- La deuxième ontologie est une ontologie du domaine. Les auteurs utilisent une ontologie pré-existante décrivant l'ensemble des problèmes cliniques et les recommandations qui leur sont liées ;
- La dernière ontologie modélise les données liées au patient.

On peut noter que ce découpage correspond au découpage des modèles de données évoqué dans la section 2. Ainsi, les contenus du GBPC sont traduits en règles logiques par les praticiens à l'aide d'une syntaxe spéciale. Elles peuvent ensuite être combinées aux connaissances contenues dans l'ontologie du domaine pour enrichir les recommandations initiales. Malheureusement, le *framework* ne propose pas de mécanisme de modélisation suivant l'évolution de l'état du patient.

Plusieurs approches proposent des modèles de GBPC fondés sur quelques primitives classiques, complétés par la création de règles SWRL contenant les connaissances décisionnelles. C'est notamment le cas de Argüello *et al.* [2009] qui proposent un alignement intéressant des concepts de l'ontologie du GBPI avec des concepts UMLS. L'approche proposée par Ceccarelli *et al.* [2008], utilisant également SWRL, considère les GBPC comme un ensemble de recommandations et utilise les primitives classiques des GBPI *action* et *decision* pour les décrire. Les résultats sont ensuite alignés avec la base de données d'un système d'information médicale afin d'analyser des données patient.

Enfin, Ongenae *et al.* [2010] proposent un travail intéressant se concentrant sur la modélisation des graphes orientés régulièrement exploités par les GBPC. À partir de dessins de graphes effectués dans une extension du logiciel Eclipse, ils en proposent une traduction semi-automatique dans un langage de règles spécifique nommé *Drools*. Ce langage peut être exécuté par un moteur spécifique, intégrant également des règles SWRL contenant des connaissances plus générales.

4.4 Conclusion partielle

De nombreux formats sont apparus au cours des quinze dernières années pour faciliter l'acquisition et l'exécution automatique des GBPI. Toutefois, aucun ne s'est imposé comme un standard et la plupart souffrent d'un manque d'implémentation dans le contexte clinique. Ce problème peut être expliqué de différentes manières. Un point de vue serait de dire que chaque format possède ses propres forces, liées aux problématiques ayant guidé son développement. Aucun ne semble assez général pour couvrir l'ensemble des problématiques liées à l'acquisition des connaissances décisionnelles. Une autre critique consiste à évoquer la variété des techniques employées, souvent *ad-hoc*. Les standards ne sont généralement pas considérés, tant au niveau de l'acquisition des connaissances que de l'exécution des GBPI. Ce qui conduit à des problèmes de portabilité des formats, difficilement intégrables à des systèmes d'information médicaux existants.

En parallèle, l'emploi des technologies du Web sémantique a été tenté dans plusieurs travaux pour proposer un modèle formel pour les GBPI. Plusieurs études ont montré la faisabilité de l'approche, et quelques propositions ont été faites. Cependant, les modèles sont liés à une application et aucun ne semble assez générique pour être utilisé en dehors de son contexte de développement.

Dans le prochain chapitre, nous présentons KCATOS, un *framework* dédié à la formalisation des GBPI. Pour répondre aux problèmes soulevés précédemment, il se veut modulaire et ouvert afin de pouvoir intégrer les différentes approches de formalisation. De plus, il repose sur l'emploi des standards du Web sémantique afin d'optimiser son interopérabilité.

Chapitre 5

Édition sémantique d'arbres de décision avec KCATOS

Sommaire

De tous les formalismes proposés pour les GBPI, aucun ne s'est imposé comme un standard à la communauté. En effet, la plupart s'avèrent complexes à appréhender ou ne disposent pas d'outils suffisamment ouverts pour être intégrés à des solutions complètes pour l'édition collaborative des GBPC. KcatoS propose de combler ce manque en s'appuyant sur les technologies du Web sémantique. Leur emploi permettra d'ouvrir les GBPI à de nouveaux standards et ainsi de bénéficier de l'apport d'outils génériques tels que les moteurs d'inférences et de ressources liées aux domaines telles que les RTO bio-médicales.

Ce chapitre présente le *framework* KcatoS. La section 5.1 présente le langage graphique construit empiriquement à partir des GBPC d'Oncolor. Les sections suivantes détaillent les méthodes d'export (section 5.2) et les extensions du langage (section 5.3). Ces dernières sont disponibles dans l'éditeur graphique collaboratif présenté dans la section 5.4. Finalement, deux évaluations sont proposées dans la section 5.5 : la première repose sur une comparaison avec les formats des GBPI présentés dans le chapitre précédent tandis que la seconde juge l'adéquation de l'outil par rapport aux arbres précédemment édités par Oncolor. Ce travail a donné lieu à des publications dans des *workshops* [Meilender *et al.*, 2011a, 2012c] et à la rédaction d'un rapport technique décrivant l'éditeur d'arbres [Meilender, 2012a].

5.1 Un langage graphique pour l'édition des GBPI

Pour créer des GBPI à partir d'arbres de décision, la définition d'un langage graphique est nécessaire. Ce langage doit être assez simple de manière à pouvoir être utilisé par des non experts, tout en conservant un degré d'expressivité suffisant pour formaliser l'ensemble des connaissances exprimées. Il doit également définir des règles syntaxiques précises afin de permettre sa conversion vers des moteurs permettant son exécution. Le langage KcatoS, présenté dans cette section, tend vers ce but. Après avoir présenté le contexte de création de ce langage et les objets graphiques le constituant, sa syntaxe et sa sémantique sont décrites.

5.1.1 Présentation générale du langage KcatoS

Contexte de création. De nombreux GBPC d'Oncolor contiennent des arbres de décision sous forme d'images bitmap, dessinés avec des outils commerciaux non collaboratifs. Ces arbres contiennent des connaissances décisionnelles dans une forme graphique non formalisée. Leur édition suit une charte graphique édictée par Oncolor, définissant en particulier les figures et les couleurs à employer. Toutefois, cette charte souffre de plusieurs manques et imprécisions. En particulier, elle ne définit aucune règle de syntaxe, ce qui a conduit à la création d'arbres à géométrie variable. Dans certains cas, les arbres manquent seulement de lisibilité mais, dans des cas plus graves, ils peuvent contenir des imprécisions ou des ambiguïtés. En outre, il convient également de constater que, dans plusieurs cas, la charte n'est simplement pas respectée. Enfin, l'édition de plusieurs arbres laisse une large part aux connaissances implicites. Ces connaissances sont des connaissances médicales nécessaires à la bonne compréhension de l'arbre qui sont considérées comme étant suffisamment générales par les rédacteurs pour ne pas être détaillées dans l'arbre.

Cependant, malgré ces défauts, Oncolor tient à conserver sa charte graphique. En effet, son élaboration date d'une dizaine d'années, ce qui a créé des habitudes d'utilisation aussi bien pour les rédacteurs que pour les lecteurs. De plus, dans le cadre de collaborations inter-réseaux, l'adoption de cette charte s'est même étendue à d'autres réseaux de cancérologie du grand Est.

Le but du langage KcatoS est donc de proposer une syntaxe pour les arbres de décision. Si cette syntaxe doit être assez simple afin de ne pas alourdir le processus d'édition des GBPI, elle doit également permettre de compenser les manques constatés dans les arbres existants. En

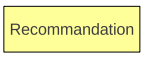
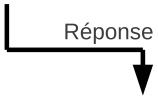
Forme	Commentaires
	Les rectangles aux angles arrondis représentent des situations médicales. Une situation médicale peut être définie comme l'état d'un patient décrit par un ensemble de variables telles que les résultats à des examens, sa physiologie, etc.
	Les recommandations sont symbolisées par des rectangles classiques. Ils contiennent l'aide à la décision proposée par le GBPI à destination des praticiens.
	Les hexagones représentent les questions dont les réponses vont permettre de décrire la situation dans le but d'obtenir la recommandation adéquate.
	Les liens sont représentés par des arcs orientés. Lorsqu'ils relient un hexagone à un autre nœud, les arcs sont typés, ce qui signifie qu'ils contiennent une réponse à la question les précédant directement.

TABLE 5.1 – Les formes et leur signification.

particulier, il faut favoriser l'explicitation de l'ensemble des connaissances médicales mises en jeu, proposer une bonne lisibilité des arbres tout en restant proche de l'existant et limiter les risques d'ambiguïté et d'incohérence. Ainsi, le langage a été construit de manière empirique après analyse des arbres existants.

Le langage KCATOS. Au final, le langage d'arbres de décision de KCATOS est une représentation graphique fondée sur un petit ensemble de formes géométriques connectées par des arcs orientés. De cette manière, les formes sont considérées comme les nœuds d'un arbre de décision. L'antécédent d'un arc est le nœud père, son image est le nœud fils. Le nœud sans père est la racine.

Du point de vue de la sémantique, chaque type de nœud a sa propre signification, comme présenté dans la table 5.1.

Cette représentation est donc inspirée directement de la charte graphique d'Oncolor. L'avantage d'utiliser ces formes géométriques est que les experts d'Oncolor les connaissent déjà. Nous cherchons donc à conserver la sémantique des graphiques d'Oncolor pour faciliter les créations et mises à jour de GBPI.

5.1.2 Syntaxe et sémantique

Syntaxe. Afin d'éviter les ambiguïtés et d'assurer la cohérence des GBPI, les règles classiques des arbres de décision sont utilisées :

- un nœud a au moins un parent, à l'exception de la racine qui n'en a pas ;
- chaque arbre a une seule racine ;
- les nœuds sont connectés par des arcs orientés ;
- un texte sur un arc représente une transition conditionnelle qui peut éventuellement prendre la forme de formules simples pouvant contenir les opérateurs booléens `ET`, `OU` et `NON`.

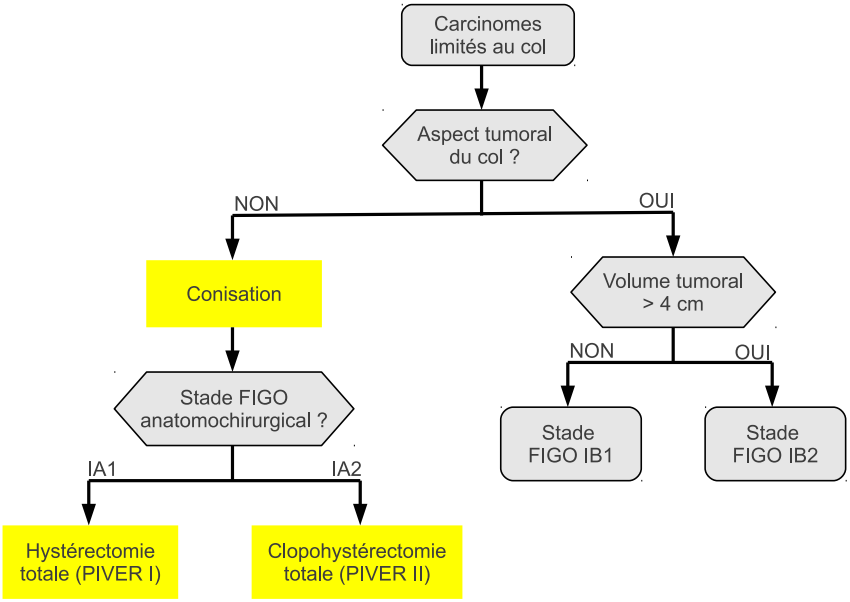


FIGURE 5.1 – Arbre syntaxiquement correct, extrait du GBPC Oncolor concernant le col de l'utérus.

Le respect de ces règles permet d'assurer que le graphe proposé est bien un arbre de décision, ce qui n'était pas le cas de tous les graphes proposés précédemment dans les GBPI d'Oncolor. De plus, quelques règles se référant aux types du langage ont été ajoutées pour garantir une sémantique correcte aux arbres :

- la racine est nécessairement une situation ;
- une question a forcément au moins une réponse ;
- chaque arc suivant une question doit contenir une réponse ;
- un texte sur un arc n'est considéré comme réponse que si cet arc suit directement une question ;
- un nœud peut avoir plusieurs parents ;
- les cycles sont interdits ;
- étant donné qu'une situation peut correspondre à plusieurs recommandations, plusieurs de ces dernières peuvent être réunies en une seule figure.

Un arbre respectant toutes ces règles est considéré comme étant syntaxiquement correct. Un tel arbre est présenté dans la figure 5.1.

Sémantique. Un arbre de décision de KcatoS peut être exporté en OWL, comme le montre la section suivante. Sa sémantique est la même que celle de la base de connaissances exportée.

5.2 Les capacités d'export

Plusieurs exports sont possibles à partir du langage KcatoS. Cette section en présente deux, chacun ayant des buts différents. La première partie, présentant un algorithme d'export, propose de formaliser les arbres de décision en OWL, montrant ainsi comment le langage s'intègre aux technologies du Web sémantique. Le second export est orienté vers l'informatique médicale et les standards des GBPI présentés dans le chapitre précédent. Son but est de montrer que le

langage peut être compatible avec des formats dédiés aux GBPI, qui sont établis et certifiés. Concrètement, les possibilités d'export vers la syntaxe Arden et GLIF sont montrées.

5.2.1 Algorithme d'export OWL

Pour améliorer la lisibilité, la section suivante utilise la logique de descriptions $\mathcal{SROIQ}(D)$ pour décrire les axiomes de la base de connaissances créée. Comme évoqué dans le chapitre 1, $\mathcal{SROIQ}(D)$ est équivalente au langage du Web sémantique OWL DL 2. En pratique seul une partie de cette LD est exploitée par l'algorithme.

5.2.1.1 Vue générale

S'appuyant sur une formalisation proposée par d'Aquin [2005], l'algorithme d'export de KcatoS se fonde sur deux classes : `Situation` et `Recommandation`. La première est relative au patient et la deuxième est relative aux décisions. Ces deux classes sont liées de la façon suivante :

$$\texttt{Situation} \sqsubseteq \exists \texttt{aPourRecommandation}.\texttt{Recommandation}$$

Cela signifie que pour toute situation σ ($\sigma \in \texttt{Situation}^{\mathcal{I}}$) il existe une recommandation ϱ ($\varrho \in \texttt{Recommandation}^{\mathcal{I}}$) qui est associée à σ ($(\sigma, \varrho) \in \texttt{aPourRecommandation}^{\mathcal{I}}$). Cette propriété `aPourRecommandation` lie une situation à une recommandation : elle a la classe `Situation` pour domaine et la classe `Recommandation` pour co-domaine.

Des sous-classes de `Situation` et `Recommandation` sont définies lors du processus d'export. Par exemple, considérons un patient qui a une céphalée et dont la recommandation est de prendre une aspirine. La classe `PatientAvecCéphalée`, sous-classe de `Situation` et la classe `PrescriptionDAspirine`, sous-classe de `Recommandation`, peuvent être introduites de la façon suivante :

$$\texttt{PatientAvecCéphalée} \equiv \texttt{Situation} \sqcap \exists \texttt{aPourSymptôme}.\texttt{SymptômeDeCéphalée}$$

$$\texttt{PrescriptionDAspirine} \sqsubseteq \texttt{Recommandation}$$

Puis, la formule suivante formalise l'assertion (médicalement discutable, mais c'est pour l'exemple) « Pour chaque patient avec une céphalée, une prescription d'aspirine est recommandée. » :

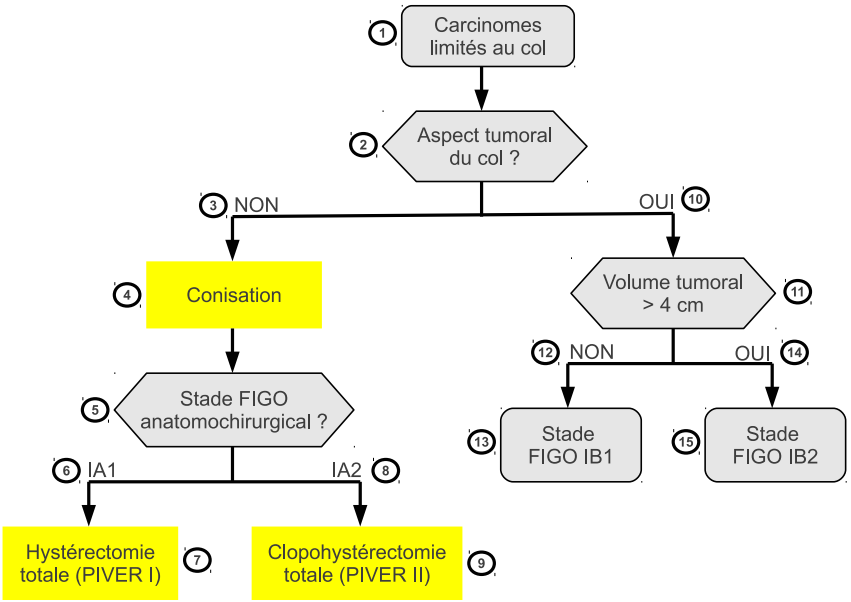
$$\texttt{PatientAvecCéphalée} \sqsubseteq \exists \texttt{aPourRecommandation}.\texttt{PrescriptionDAspirine}$$

5.2.1.2 Règles de traduction

Un arbre de décision de KcatoS est lu selon un parcours en profondeur. Chaque nœud est transformé en utilisant des règles prenant en compte ses ancêtres. Ces règles sont expliquées ci-dessous. Un exemple complet d'export d'un arbre de décision vers OWL est montré à la figure 5.2.

Situations. Une forme de situation permet de créer une classe `Sit_Y` comme sous-classe de `Sit_X`, où `Sit_X` est la plus proche sous-classe de `Situation` (*i.e.*, la sous-classe associée au nœud parent ou à l'arc parent).

$$\texttt{Sit\_Y} \sqsubseteq \texttt{Sit\_X}$$



(a) Arbre extrait du GBPI Oncolor du cancer du col de l'utérus limité au col.

1. $\texttt{SitCLC} \sqsubseteq \texttt{Situation}$
2. $\texttt{aATC}$: propriété de type de données fonctionnelle
   domaine : $\texttt{SitCLC}$
   co-domaine : $\texttt{booléen}$
3. $\texttt{SitATC\_Faux} \equiv \texttt{SitCLC} \sqcap \ni \texttt{aATC.faux}$
4. $\texttt{SitATC\_Faux} \sqsubseteq \exists \texttt{aPourRecommandation.Conisation}$
5. $\texttt{aPourStadeFigo}$: propriété object fonctionnelle
   domaine : $\texttt{SitATC\_Faux}$
   co-domaine : $\texttt{RepStadeFigo}$
6. $\texttt{RepStadeFigo(IA1)}$
   $\texttt{SitStadeFigoIA1} \equiv \texttt{SitATC\_Faux} \sqcap \ni \texttt{aPourStadeFigo.IA1}$
7. $\texttt{SitStadeFigoIA1} \sqsubseteq \exists \texttt{aPourRecommandation.PiverI}$
8. $\texttt{RepStadeFigo(IA2)}$
   $\texttt{SitStadeFigoIA2} \equiv \texttt{SitATC\_Faux} \sqcap \ni \texttt{aPourStadeFigo.IA2}$
9. $\texttt{SitStadeFigoIA2} \sqsubseteq \exists \texttt{aPourRecommandation.PiverII}$
10. $\texttt{SitATC\_Vrai} \equiv \texttt{SitCLC} \sqcap \ni \texttt{aATC.vrai}$
11. $\texttt{aVolumeSup4}$: propriété de type de données fonctionnelle
    domaine : $\texttt{SitATC\_Vrai}$
    co-domaine : $\texttt{booléen}$
12. $\texttt{SitVolumeSup4\_Faux} \equiv \texttt{SitATC\_Vrai} \sqcap \ni \texttt{aVolumeSup4.faux}$
13. $\texttt{SitFigoIB1} \sqsubseteq \texttt{SitVolumeSup4\_Faux}$
14. $\texttt{SitVolumeSup4\_Vrai} \equiv \texttt{SitATC\_Vrai} \sqcap \ni \texttt{aVolumeSup4.vrai}$
15. $\texttt{SitFigoIB2} \sqsubseteq \texttt{SitVolumeSup4\_Vrai}$

(b) Traduction en $\mathcal{SROIQ}(D)$ de l'arbre précédent.

FIGURE 5.2 – Un extrait du GBPI Oncolor du cancer du col de l'utérus limité au col (a) et sa traduction en OWL (b).

Recommandations. Une forme de recommandation indique qu'une classe de situation `Sit_X` est liée à une recommandation `Reco1` par la propriété `aPourRecommandation`.

$$\texttt{Sit\_X} \sqsubseteq \exists \texttt{aPourRecommandation}.\texttt{Reco1}$$

Questions. Une forme de question introduit une nouvelle propriété fonctionnelle, `aPourRéponse`, ayant `Sit_X`, la plus proche sous-classe de situation, comme domaine. Si les réponses sont `"vrai"` et `"faux"` ou `"oui"` et `"non"`, `aPourRéponse` est une propriété de type de données avec un co-domaine booléen. Sinon, c'est une propriété objet ayant une nouvelle sous-classe, `RéponseÀQuestion`, comme co-domaine.

`aPourRéponse` : propriété fonctionnelle
domaine : `Sit_X`
co-domaine : `booléen`
ou `RéponseàQuestion`

Arcs. Comme cela a été écrit dans la section 5.1, un arc contient une réponse `RÉPONSE` à la question `aPourRéponse` qui le suit directement. Il introduit une nouvelle sous-classe de `Sit_X`, en spécifiant sa valeur de propriété.

$$\texttt{Sit\_Y} \equiv \texttt{Sit\_X} \sqcap \ni \texttt{aPourRéponse}.\texttt{RÉPONSE}$$

5.2.2 Export vers les standards du HL7

5.2.2.1 Transformation en syntaxe Arden

Comme expliqué dans le chapitre précédent, à chaque MLM correspond une seule recommandation. Les arbres KcatoS contenant plusieurs recommandations, à chaque arbre KcatoS va donc correspondre plusieurs MLM. De manière générale, il convient donc d'isoler chaque chemin conduisant à une recommandation ou à une feuille, pour ensuite le décrire dans le MLM dédié. Les liens entre les différents chemins ainsi isolés peuvent être conservés par l'intermédiaire de variables de type *event* qui correspondent à travers les *slots Evoke* et *Action*.

Dans la suite, nous allons détailler l'export des *slots* les plus intéressants. Il sera illustré par des fragments d'ArdenML, construit à partir de l'arbre de décision de la Figure 5.2a, plus particulièrement le chemin menant à la recommandation *Conisation*, qui réduit l'extraction aux nœuds 1,2 et 4, et à leurs arcs. L'export est réalisé à partir du fichier XML propre à KcatoS conservant toutes les informations textuelles et graphiques.

Dans un souci de lisibilité, les modules présentés utiliseront une syntaxe XML (dite *ArdenML*) proposée par Kim *et al.* [2008].

Construction des *slots Maintenance* et *Library*. Cette partie est informative et ne nécessite pas de traitement spécial. Un simple formulaire HTML peut suffire à recueillir son contenu. Dans le cadre d'un système opérationnel de gestion des GBPI, il conviendrait de relier cette partie aux modules de gestion des utilisateurs.

De la même manière, *Library* ne peut être rempli automatiquement. Toutefois, les informations qu'il contient sont importantes à conserver car elles seront réutilisées fréquemment dans le cadre des mises à jour. Ces *slots* sont présentés dans la figure 5.3.

```
<Maintenance>
        <Title>Test Conisation</Title>
        <MLMName>Test_Conisation.xlm</MLMName>
        <Arden>Version 2.7</Arden>
        <Version>1.00</Version>
        <Institution>
            <Name_of_Institution>Projet Kasimir</Name_of_Institution>
        </Institution>
        <Author>...</Author>
        <Specialist />
        <Date>2011-02-19</Date>
        <Validation>test</Validation>
</Maintenance>
<Library>
        <Purpose>
         D efinir les cas de recommandation d'une conisation
        </Purpose>
        <Explanation>
            Si le col a un aspect tumoral, alors une conisation est recommandée.
        </Explanation>
        <Keywords>
            <Keyword>carcinomes</Keyword>
            <Keyword>conisation</Keyword>
            <Keyword>col</Keyword>
        </Keywords>
</Library>
```

FIGURE 5.3 – Export vers les *slots Maintenance* et *Library*.

Construction du *slot Data*. Ce *slot* regroupe les données nécessaires à l'exécution du MLM. Dans le cadre de notre exemple, les données sont de deux types : (1) les données liées aux tests logiques et (2) les données liées aux appels de modules externes. À chaque question présente dans le chemin, une variable est associée, dont le type sera un booléen si la réponse est `"vrai"` ou `"faux"`, `"oui"` ou `"non"`, une chaîne de caractères sinon. Le texte présenté au praticien pour la saisie de la valeur est directement issu des questions présentes dans le chemin. Dans le cas des modules externes, on déclare une variable booléenne de type événement afin de gérer le déclenchement de modules externes. Cette variable n'est pas nécessaire si la recommandation est une feuille dans l'arbre de décision. La figure 5.4 présente le code XML correspondant à ce *slot* pour l'exemple traité.

Construction du *slot Evoke*. Dans cette partie, ArdenML (voir figure 5.5) se contente de préciser quel est l'événement qui déclenchera l'utilisation du MLM. Cet événement prend la forme d'une variable booléenne : son passage à vrai active le module. Cette variable est liée à la forme de situation 1 de l'arbre de décision, dans le sens où elle sera utilisée par chaque MLM dont le chemin qu'il modélise contient cette forme.

Construction des *slot Logic* et *Action*. Présenté dans la figure 5.6, le *slot Logic* est construit à partir des questions présentes dans le chemin. Les tests sont directement importés depuis les formes et les réponses depuis les arcs. Ils mettent en forme les conditions en se servant des données utilisateur saisies dans la partie *data* et proposent les conclusions *true* ou *false*, activant le *slot Action* dans le cas d'une réponse positive. Ce dernier, présenté dans la figure 5.7, indique l'action à effectuer le cas échéant.

```
<Data>
    <Read>
        <Identifier var="Aspect_Tumoral" otype="boolean"/>
        <Assigned>
            <Latest>
                <Mapping>
                    <Contents>"Aspect Tumoral du col ?"</Contents>
                    <XForms>
                        <select1>
                            <label>Aspect_Tumoral</label>
                            <item><value>true</value></item>
                            <item><value>false</value></item>
                        </select1>
                    </XForms>
                </Mapping>
            </Latest>
        </Assigned>
    </Read>
    <Event>
        <Identifier var="Carcinommes_Lim_Au_Col_Conisation" otype="boolean"/>
        <Assigned>
            <Mapping>
                <Contents>
                    "Carcinommes_Lim_Au_Col_Conisation_Appel"
                </Contents>
            </Mapping>
        </Assigned>
    </Event>
</Data>
```

FIGURE 5.4 – Export vers le *slot Data*.

```
<Evoke>
    <Identifier var="Carcinomes_Lim_Au_Col_Appel" otype="boolean"/>
</Evoke>
```

FIGURE 5.5 – Export vers le *slot Evoke*.

```
<Logic>
    <If>
        <Condition>
            <Identifier var="Aspect_Tumoral" otype="boolean"/>
        </Condition>
        <Then>
            <Conclude><Value otype="boolean">true</Value></Conclude>
        </Then>
        <Else>
            <Conclude><Value otype="boolean">false</Value></Conclude>
        </Else>
    </If>
</Logic>
```

FIGURE 5.6 – Export vers le *slot Logic*.

```
<Action>
    <Write>
        <Value otype="string">Recommandation : Conisation.</Value>
    </Write>
    <Assignment>
        <Identifier var="Carcinommes_Lim_Au_Col_Conisation" otype="boolean"/>
        <Assigned><Value otype="boolean">true</Value></Assigned>
    </Assignment>
</Action>
```

FIGURE 5.7 – Export vers le *slot Action*.

5.2.2.2 Étude d'un export vers GLIF

Si l'export des grands référentiels vers la syntaxe Arden est possible, le résultat est peu lisible et son exploitation nécessiterait la conception d'un moteur spécifique. Comme évoqué dans le chapitre précédent, GLIF est un autre standard du HL7, conçu comme un *framework* destiné à la modélisation de référentiels plus complexes.

GLIF est bien plus expressif que la syntaxe Arden, mais moins utilisé dans les institutions médicales, notamment à cause de son manque d'implémentation et de la complexité de mise en place de sa méthode de conception. Toutefois, les deux standards présentent plusieurs similarités :

- le langage GEL utilisé pour exprimer des expressions logiques est largement inspiré de la syntaxe Arden,
- le mécanisme d'appel d'action (*slot action*) est le même,
- GLIF propose un mécanisme de macro nommé « MLMs macro » utilisant des *slots logic* et *action* à la manière de la syntaxe Arden.

Dans leur étude de rapprochement de ces standards, Peleg *et al.* [2001] traitent de ces compatibilités et, après avoir montré la complémentarité des approches, évoquent le développement d'un langage commun entre la syntaxe Arden et les macros MLM. Ainsi, si KcatoS est compatible avec la syntaxe Arden, il le sera avec ce nouveau langage, lui même compatible avec GLIF. L'apport de KcatoS au *framework* serait double : il fournirait une interface graphique l'implémentant et permettrait de simplifier la méthode de conception en prenant en charge le rôle de l'ingénieur des connaissances lors de la deuxième phase. Mais, pour pouvoir utiliser pleinement les possibilités de GLIF, KcatoS doit améliorer son expressivité, notamment en prenant en compte les options de traitement qui ne sont pas régies par une équation logique mais par un choix du praticien et en intégrant la notion d'adaptation du référentiel aux contraintes locales. On peut noter que cette première contrainte est résolue dans la section suivante.

5.3 Extensions du langage

La simplicité du langage KcatoS le rend modulaire et permet l'intégration d'extensions. Ces extensions sont construites en fonction des besoins rencontrés lors de l'édition des GBPI. Leur but est donc d'étendre le langage de manière à améliorer son expressivité. Dans le cadre du *framework*, elles doivent préciser deux niveaux : le niveaux graphique et le niveau de formalisation des connaissances. La définition du niveau graphique consiste à présenter la forme géométrique utilisée en contexte et à donner les éléments nécessaires à la traduction des connaissances de manière à permettre la formalisation complète des arbres.

Ainsi, plusieurs extensions ont déjà été définies pour correspondre aux besoins des GBPI d'Oncolor. Celles-ci sont décrites dans les paragraphes suivants.

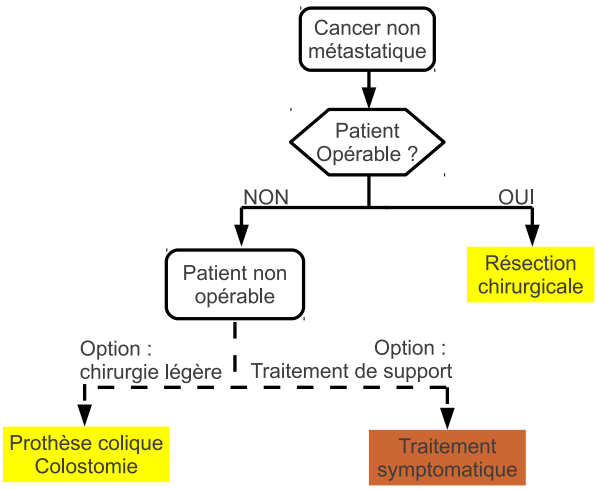


FIGURE 5.8 – Extrait d'arbre de décision adapté du GBPI Oncolor concernant le côlon.

Extension *Option*. Un nouveau type d'arcs est défini dans cette extension, nommé « Option ». Graphiquement, il prend la forme d'une flèche tracée en pointillés, comportant un texte. Ces options correspondent à des choix thérapeutiques qui peuvent être appliqués à une situation. D'un point de vue médical, le choix de l'option ne dépend pas de critères formalisables mais intégralement des préférences de l'utilisateur. Ainsi, les options permettent de proposer un certain nombre de pratiques possibles au thérapeute et de guider le processus d'exécution en fonction du choix de celui-ci. Un exemple d'utilisation de cas d'utilisation est montré à la figure 5.8.

Ainsi, les options sont des formes particulières d'arcs. Mais, contrairement aux arcs classiques décrits plus tôt, ils définissent une situation en renseignant une propriété `aPourOption` et créent des instances de la classe `OptionThérapeutique`.

$$\texttt{Sit\_Y} \equiv \texttt{Sit\_X} \sqcap \ni \texttt{aPourOption.OPTION}$$

Extension *Catégories de recommandation*. Dans la charte graphique d'Oncolor est définie une politique de classification des recommandations. En effet, selon sa nature, chaque recommandation est classée selon 7 catégories en fonction des spécialités médicales nécessaires à son application. D'un point de vue graphique, cette classification est mise en valeur par la couleur utilisée pour remplir la figure de recommandation. Ainsi, par exemple, les actes chirurgicaux apparaissent en jaune et les traitements à base de rayons apparaissent en rose. Ce code de couleur est intégralement décrit dans la charte graphique d'Oncolor. Un exemple d'arbre utilisant ce code est montré dans la figure 5.8.

Lors de la création de la base de connaissances, cette distinction des recommandations se décline par la création de sous-classes à la classe `Recommandation`.

Extension *Lien*. Cette extension permet de définir des liens entre les arbres, prenant graphiquement la forme d'ellipses. Elles permettent ainsi de construire des arbres volumineux divisés en plusieurs parties. Elles peuvent également désigner des liens vers des recommandations prédéfinies fréquentes, telles que la préconisation d'une réunion de concertation pluridisciplinaire. L'emploi de cette extension est montré dans la figure 5.9.

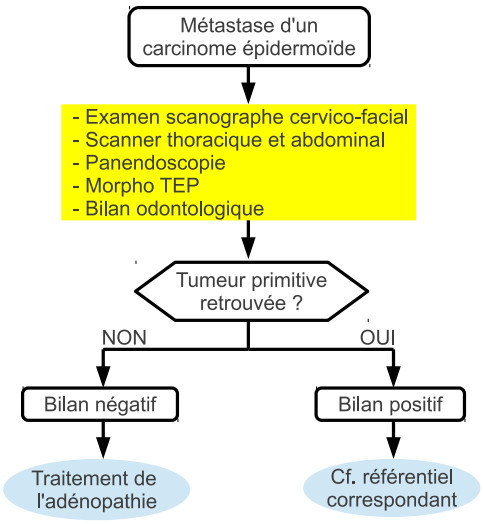


FIGURE 5.9 – Extrait d'arbre du GBPI Oncolor concernant le côlon.

Les liens sous forme d'ellipses peuvent apparaître dans les arbres sous forme de racines et de feuilles. Donc, une situation `Sit_X` décrite dans un premier arbre est équivalente à une situation `Sit_Y` vers laquelle elle pointe dans un deuxième arbre :

$$\texttt{Sit\_Y} \equiv \texttt{Sit\_X}$$

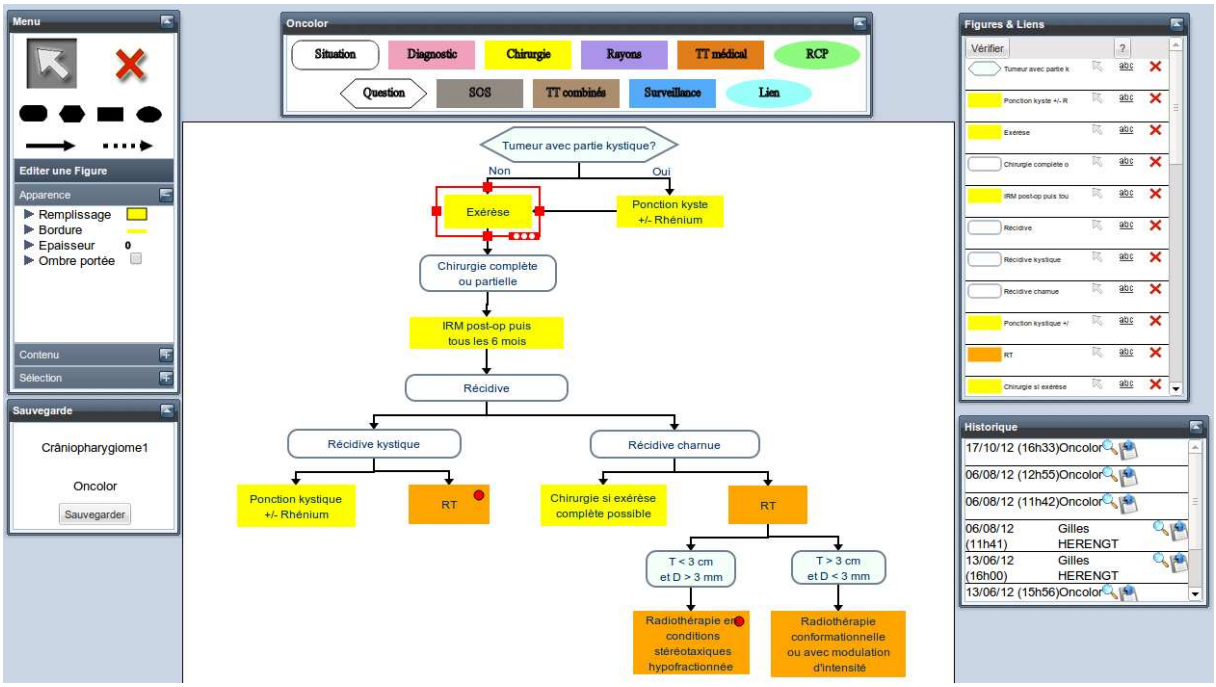
5.4 L'éditeur KCATOS

Afin de permettre l'utilisation du langage KCATOS dans le contexte d'application de ce travail, un éditeur graphique d'arbre de décision a été proposé. Le *framework* KCATOS propose une application en ligne permettant de dessiner de manière collaborative les arbres, de les stocker et de les convertir vers des langages du Web sémantique. Cette section présente le *framework* d'un point de vue technique et ses principales fonctionnalités, puis montre ses possibilités d'intégration à une solution plus globale pour l'édition des GBPI. Enfin, est évoquée de quelle manière il peut devenir un élément d'un système d'aide à la décision s'il est couplé à d'autres outils du projet KASIMIR tels que KOWL et EDHIBOU.

5.4.1 Présentation de l'éditeur

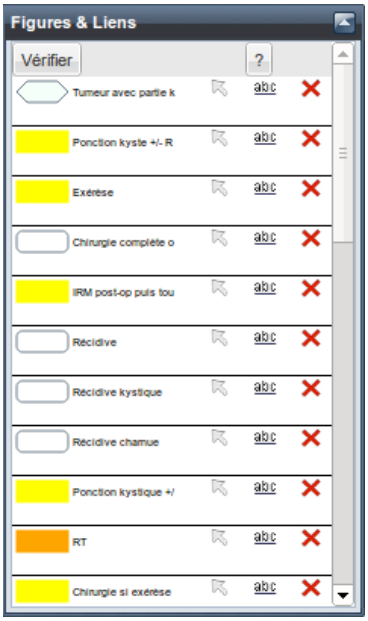
L'éditeur d'arbres de décision KCATOS est une application Web créée avec Google Web Toolkit (GWT) [Google, 2013b] qui permet de créer des applications Ajax complexes. Il utilise également quelques API complémentaires dédiées à GWT pour gérer l'interface, en particulier pour le dessin de figures. Les capacités de dessin reposent sur l'emploi des technologies SVG et JavaScript. Comme Internet Explorer ne supporte pas SVG sans greffon jusqu'à la neuvième version, KCATOS propose une version exploitant le langage graphique SVML.

En tant qu'application Web, KCATOS est ouvert au travail collaboratif et à l'utilisation de services Web. Il peut être intégré à la plupart des systèmes de gestion de contenu (CMS) : des tests ont déjà été accomplis avec le CMS MediaWiki dans le cadre d'OncoLogiK. Cette intégration est d'ailleurs présentée dans la section 5.4.2. Actuellement, KCATOS a été testé avec succès avec les navigateurs Mozilla, Opera, Chrome et Internet Explorer.



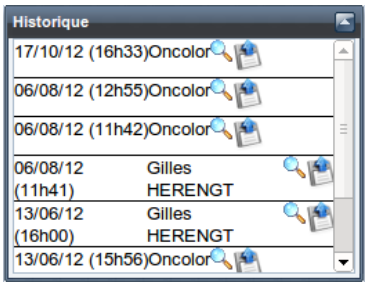
(a) Interface de l'éditeur d'arbres KcatoS.

(b) Barre de menu.



(c) Barre « Figures et Liens ».

(d) Barre d'informations.



(e) Barre d'historique.

(f) Édition dans le langage KcatoS.

FIGURE 5.10 – Interface de l'éditeur d'arbres KcatoS et détail de ses composants.

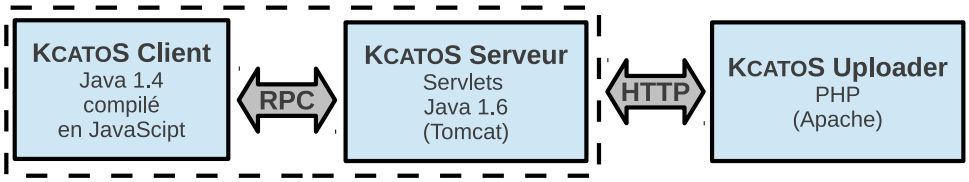


Figure 5.11 – Architecture générale de KcatoS.

Une capture d'écran d'une version autonome est montrée dans la figure 5.10a. Il est disponible sous une licence libre (LGPL).

Fonctionnalités. Le *framework* KcatoS propose de nombreuses fonctionnalités aux utilisateurs. En particulier, plusieurs exports sont possibles :

- vers XML, dans un format dédié,
- pour la création d'images, en proposant des exports vers des formats bitmap (JPG, PNG) et un format vectoriel (SVG),
- pour la création de bases de connaissances en OWL, suivant l'algorithme d'export présenté dans la section précédente, en exploitant OWL API [Horridge et Bechhofer, 2011].

KcatoS propose en outre des outils pour le travail collaboratif. Ainsi, chaque arbre stocké sur le serveur dispose de son historique de création, ce qui permet de revenir à une version antérieure à la version courante. De plus, afin de garantir que la syntaxe du langage est bien respectée, un module peut être utilisé pour vérifier si les arbres édités respectent les règles syntaxiques définies précédemment. Les extensions du langage KcatoS présentées dans la section précédente sont toutes prises en charge par l'éditeur.

Architecture. Comme le montre la figure 5.11, l'architecture de KcatoS se divise en trois parties :

- KcatoS *client* définit les composants graphiques de l'interface à laquelle accède l'utilisateur. Cette partie génère l'interface graphique en Javascript grâce au compilateur de GWT et gère les interactions avec l'utilisateur. Elle permet l'édition des arbres de décision grâce à SVG (ou SVML, le cas échéant).
- KcatoS *serveur* contient les fonctionnalités de KcatoS exécutées du coté serveur, en particulier la création des fichiers et la communication avec la troisième partie.
- KcatoS *uploader* gère les fonctionnalités liées aux fichiers créés et la communication avec des applications externes. Cette partie est développée en PHP et installée sur un serveur distant. Elle communique avec le reste de l'application par requêtes HTTP.

Les deux premières parties sont exclusivement gérées par GWT tandis que la troisième est indépendante. Elle peut être implémentée de manière différente selon le contexte d'application et les fonctionnalités désirées.

Lors de sa programmation, le *framework* a été conçu comme suffisamment générique pour être intégré à différents projets, même s'ils sont extérieurs au domaine de l'oncologie. Ce but a été atteint puisque KcatoS est également utilisé dans le projet Taaable, dont un des objectifs est la formalisation et l'adaptation des recettes de cuisine. Dans ce cadre, KcatoS est utilisé pour dessiner une représentation de l'exécution de la recette [Dufour-Lussier *et al.*, 2012].

Description de l'interface. L'interface, présentée dans la figure 5.10a, est composée d'une page principale, permettant la création des arbres, et de cinq barres d'outils :

- La barre `Menu` (figure 5.10b) est divisée en 2 parties :
  - la partie haute contient l'accès aux commandes principales permettant de sélectionner le mode d'édition (création d'une figure, sélection/modification, suppression),
  - la partie basse contient des commandes contextuelles, relatives au mode d'édition sélectionné dans la partie haute (par exemple choix de la couleur, etc.).
- La barre `Sauvegarde` (figure 5.10d) affiche des informations concernant l'arbre en cours d'édition (titre, auteur) et le bouton de sauvegarde.
- La barre `Figures & Liens` (figure 5.10c) présente l'ensemble des figures et des liens édités, et propose des liens pour l'édition et la suppression de ceux-ci. Elle contient également l'accès aux fonctions de vérification de la syntaxe d'un arbre.
- La barre `Historique` (figure 5.10e) pointe vers les versions précédentes de l'arbre édité. Elle affiche diverses informations le concernant (auteur, date) et fournit des liens permettant de prévisualiser et de charger la version voulue.
- La barre `Oncolor` (figure 5.10f) contient des squelettes prédéfinis de figures utiles à l'élaboration des arbres par Oncolor. Elle correspond à la charte graphique établie courant 2011, et son code la rend facilement adaptable en cas de modification.

5.4.2 Intégration à OncoLogiK

Comme évoqué précédemment, le *framework* KcatoS a été intégré à OncoLogiK afin de fournir à Oncolor une solution complète pour l'édition des GBPI. La figure 5.12 montre l'exemple d'un arbre édité dans le wiki sémantique.

Une communication bidirectionnelle a été établie entre OncoLogiK et KcatoS. La figure 5.13 en décrit le fonctionnement. Outre les requêtes HTTP fournissant un support à cette communication, elle repose principalement sur les échanges entre la partie KcatoS *uploader* et l'API de SMW. L'affichage des arbres repose sur l'emploi de modèles spécialement adaptés dans OncoLogiK. Ainsi, chaque arbre est représenté par une page dans le wiki, dont l'intégration à la page du GBPI se fait de la même manière qu'une image classique.

Grâce à cette communication, OncoLogiK peut utiliser ainsi les capacités d'export de l'éditeur pour l'affichage et l'exploitation des arbres de décision. Pour sa part, KcatoS propose des outils en lien avec le wiki lors de la création des arbres en proposant en particulier l'édition de liens hypertextes vers les contenus d'OncoLogiK.

5.4.3 Vers un système d'aide à la décision

L'export des arbres vers OWL permet d'envisager leur utilisation dans le contexte d'applications du Web sémantique. En particulier, le projet Kasimir propose des outils pour interroger ces bases de connaissances. Du point de vue de l'informatique médicale, ces outils permettent l'exécution des GBPI, ce qui revient à proposer les bases d'un système d'aide à la décision. Ainsi, cette section montre comment peut être interrogée la base de connaissances à deux niveaux :

- le premier niveau montre l'exploitation des connaissances par requêtes SPARQL avec le serveur de connaissances KOWL, permettant l'utilisation d'un moteur d'inférences,
- le deuxième niveau, plus graphique et simple d'accès, propose une interface au système par l'intermédiaire du *framework* EdHibou.

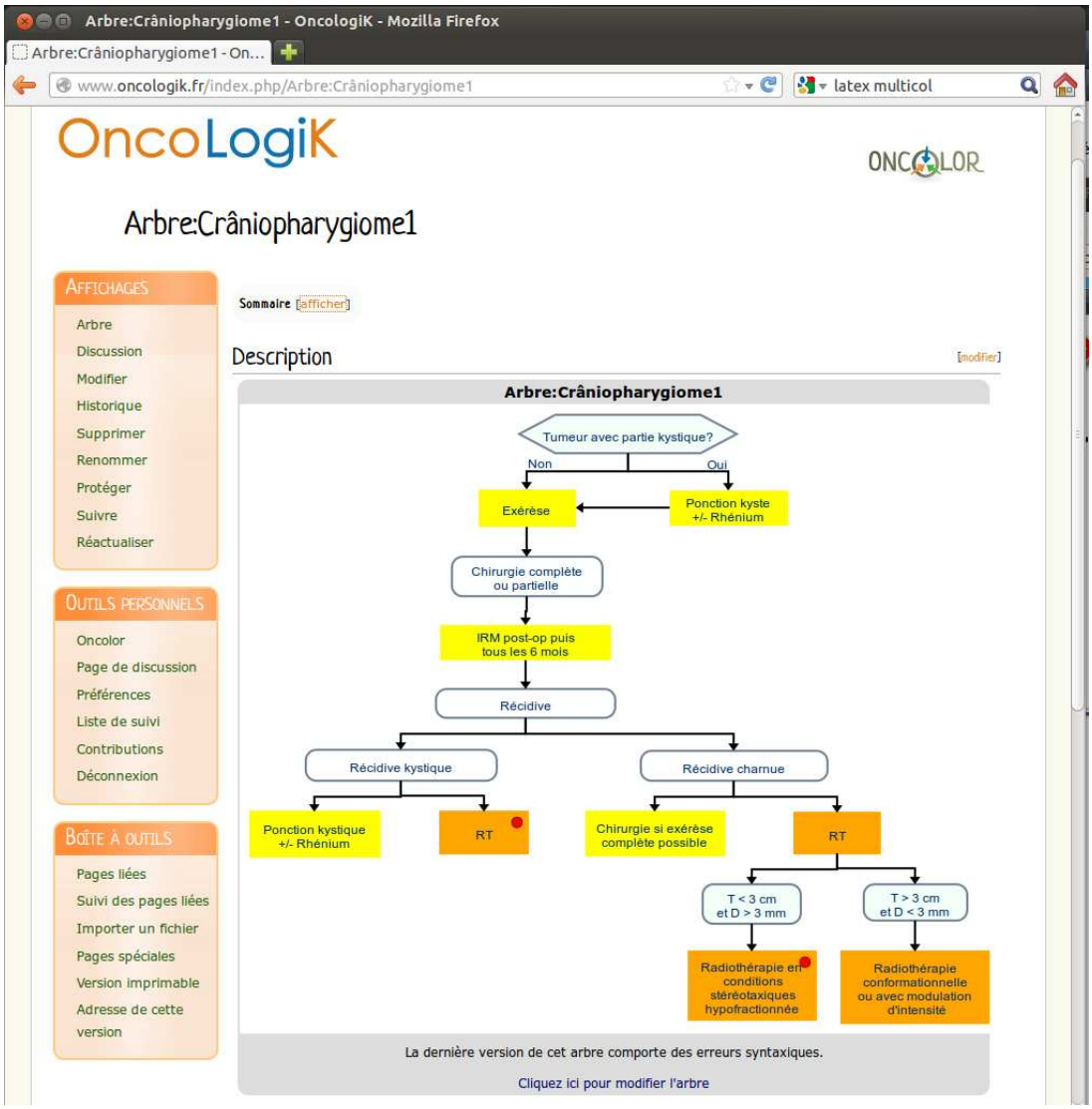


FIGURE 5.12 – Page d'OncoLogiK présentant un arbre édité avec KcatoS.

Exploitation des connaissances avec Kowl. Développé dans le cadre du projet Kasimir, Kowl est un serveur de connaissance pour le Web sémantique. Le temps d'une session, il permet le peuplement d'une ontologie OWL par l'ajout d'instances et d'assertions. Il permet également l'interrogation de cette ontologie par l'utilisation du langage de requêtes SPARQL. Kowl supporte également le raisonnement en prenant en charge la communication avec un moteur d'inférences. Dans le cadre de ce travail, le moteur d'inférences utilisé est Pellet [Sirin *et al.*, 2007a]. Un exemple d'application est donné à la figure 5.14.

KcatoS et EdHibou. Développé dans le cadre de la boîte à outils KatexOwl, EdHibou est un *framework* pouvant être implémenté en tant que service Web. Son but est de fournir une interface utilisateur pour Kasimir qui permette la description d'une situation médicale et fournisse la recommandation adaptée. Techniquement, le principe de l'application est de créer une instance dans une ontologie OWL et de considérer les propriétés liées comme des questions

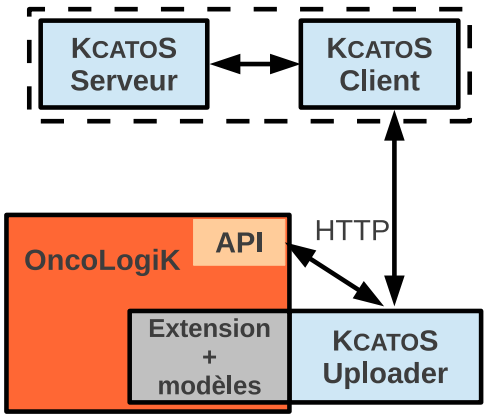


FIGURE 5.13 – Communication entre KCATOS et OncoLogiK.

auxquelles l'utilisateur peut répondre via l'interface. EDHIBOU propose des vues annexes paramétrables afin de suivre l'évolution d'éléments dans l'ontologie en fonction des choix effectués par l'utilisateur, par exemple les classes d'une instance.

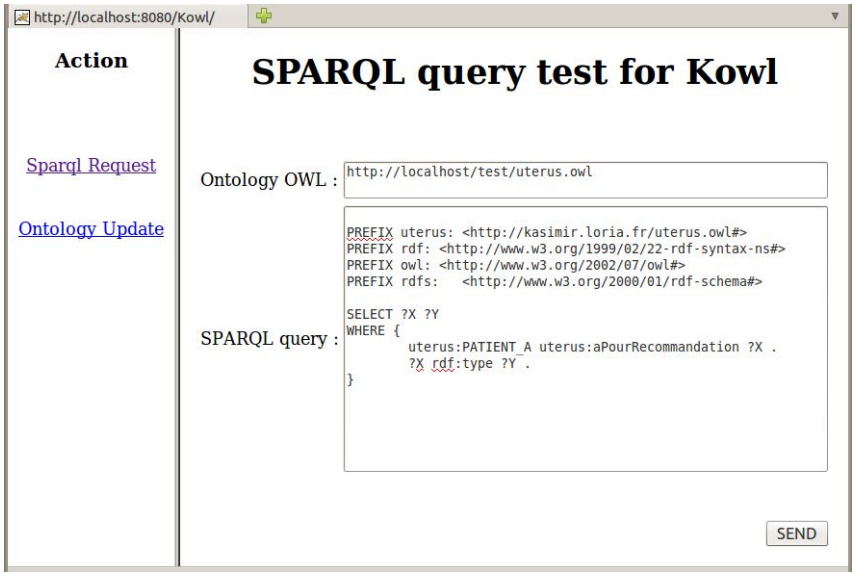
Dans le contexte des GBPI, EDHIBOU utilise les ontologies OWL créées par KCATOS. Une instance de `Situation` et une instance de `Recommandation` sont créées et les propriétés reliées à cette situation sont proposées sous forme de question. Une vue est paramétrée afin de suivre l'évolution de l'instance `Recommandation` et ainsi afficher les recommandations adéquates. Combiné à KCATOS, il crée un formulaire correspondant à l'arbre de décision construit par l'utilisateur comme on peut le voir dans la figure 5.15. L'apport de cette combinaison d'outils du Web sémantique est de permettre à l'utilisateur de vérifier que l'interprétation faite par KCATOS correspond à ce qu'il souhaite modéliser.

5.5 Évaluation des capacités de modélisation

5.5.1 Comparaison avec les autres formats de GBPI

Modèle de données. Comme montré dans le chapitre précédent, les formats les plus récents proposent un découpage en trois modèles : un modèle de décision, un modèle du domaine et un modèle du dossier patient. Le *framework* KCATOS propose actuellement uniquement un modèle de décision. Les primitives qu'il emploie sont proches de celles employées par GLIF et SAGE par exemple. La possibilité d'export vers GLIF illustre d'ailleurs cette proximité. De même, le choix d'une représentation graphique sous forme d'arbre de décision est commun à de nombreux langages, en particulier pour leur lisibilité. Les transitions entre les états se font sous forme de propositions booléennes. Certains formats proposent d'intégrer des fonctions de calcul supplémentaires, comme dans le langage GELLO, ou des contraintes temporelles, à l'image d'Asbru et de PRO*forma*.

Toutefois, contrairement à l'ensemble des formats pré-existants, le modèle de données de KCATOS est traduisible dans le langage de représentation OWL, ce qui offre une portabilité absente de ses prédécesseurs. Cette portabilité s'exprime à deux niveaux : le premier niveau concerne l'édition des connaissances, puisque les bases de connaissances peuvent être créées avec des éditeurs OWL comme Protégé ou NeOn Toolkit, et le second niveau concerne l'exécution, possible avec tous les moteurs d'inférences compatibles avec OWL tels que par exemple Pellet.



(a) Interrogation de la base de la figure 5.2 destinée à déterminer les classes de la recommandation `MA_RECO`, où cette dernière est reliée à une situation `Patient_A` pour laquelle la propriété `aATC` a la valeur *false*.

X	Y
http ://kasimir.loria.fr/uterus.owl#MA_RECO	http ://kasimir.loria.fr/uterus.owl#Conisation
http ://kasimir.loria.fr/uterus.owl#MA_RECO	http ://www.w3.org/2002/07/owl#Thing
http ://kasimir.loria.fr/uterus.owl#MA_RECO	http ://kasimir.loria.fr/uterus.owl#Recommendation

(b) Réponse à la requête précédente indiquant que l'instance `MA_RECO` est membre des classes `owl:Thing`, `Recommandation` et `Conisation`.

FIGURE 5.14 – Requête SPARQL (a) sur KOWL et sa réponse (b) destinées à déterminer les classes d'une recommandation liée à une instance de situation définie.

De plus, la structure modulaire du langage ouvre la voie à de nouvelles extensions, dont certaines pourront égaler les avantages constatés dans les autres langages.

Actuellement, les modèles du domaine et du dossier patient ne sont pas proposés. Ces absences sont fortement liées puisque le lien entre un guide et le dossier patient est généralement pris en charge par le modèle du domaine. L'absence de modèle du domaine est due à un choix préalable destiné à favoriser l'utilisation du langage. En effet, aucune contrainte d'édition des textes contenus dans les figures n'existe, ce qui implique de ne pas imposer la présence d'objets d'un modèle de données tel que MeSH ou SNOMED. Le chapitre 6 proposant des perspectives pour le projet donne des pistes pour intégrer progressivement de nouveaux modèles.

Outils d'édition. KCATOS propose un outil d'édition d'arbres de décision, à l'image de nombreux formats. Celui-ci possède toutefois l'avantage d'être disponible en ligne et d'offrir des fonctionnalités permettant le travail collaboratif. De plus, son adossement à un wiki sémantique permet d'ouvrir la voie à de nouvelles possibilités, en particulier concernant la description du contexte d'édition d'un arbre.

Exécution. L'exécution des guides repose sur le framework EDHIBOU. Cette intégration est rendue possible par la portabilité du langage KCATOS. Contrairement aux autres formats pro-

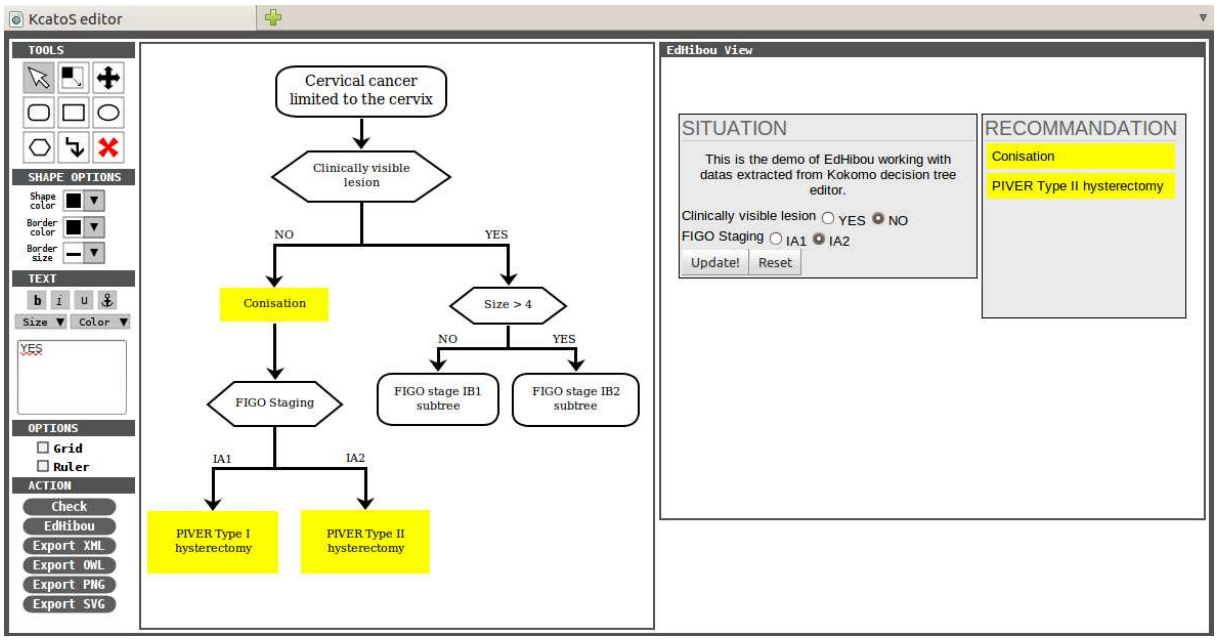


FIGURE 5.15 – KCATOS combiné à EDHIBOU.

posant des moteurs d'exécution *ad-hoc*, les bases de connaissances peuvent être consultées avec des outils reposant sur les standards du Web sémantique tels qu'OWL et SPARQL.

5.5.2 Adaptation aux besoins de représentation des GBPC d'Oncolor

Traitement des arbres existants. Plus de 600 arbres de décision illustrent déjà les GPBC d'Oncolor. Ces arbres sont stockés sous forme d'images bitmap qui devront être transformées pour pouvoir être prises en compte par KCATOS. Les arbres de décision ont évolué au fur et à mesure de leurs mises à jour et, malheureusement, les experts médicaux n'ont pas toujours pris soin de respecter ni les règles classiques des arbres de décision, ni la charte graphique. Les arbres qui en résultent sont toujours lisibles et compréhensibles par des spécialistes, mais leur formalisme est loin d'être systématiquement reconnaissable. Ce problème exclut une transformation automatique des arbres dans un format compatible avec KCATOS. Comme les GBPC sont régulièrement mis à jour, il a été décidé que l'ensemble des arbres sera progressivement redessiné avec KCATOS.

Cependant, les algorithmes de vérification de syntaxe de KCATOS n'acceptent pas les arbres de décision qui ne sont pas syntaxiquement corrects. Après avoir analysé un échantillon de 150 arbres décisionnels existants, il est apparu que seulement 44 d'entre eux pourraient être considérés comme bien formés. En étudiant les causes d'erreurs, certaines semblent fréquentes et facilement identifiables, ne nécessitant que de simples corrections. Par exemple, 42 arbres de décision ne proposaient pas de racine correcte, c'est-à-dire qu'ils ne commençaient pas par une situation. La plupart du temps, la forme a été oubliée par les rédacteurs alors qu'elle est décrite dans une partie du texte du GBPC. Dans ces cas, une figure de type `Situation` a été tout simplement ajoutée à la racine de l'arbre. Une autre erreur fréquente est l'absence de texte accompagnant une transition suivant une question. Cette erreur a été observée dans 47 arbres. En règle générale, le texte a été placé à l'intérieur d'une forme `Situation` liée à la question. La correction est assez simple : le texte doit être déplacé et la situation doit être supprimée.

Après ces corrections, 62 arbres présentent encore des erreurs nécessitant des corrections spécifiques ou plus complexes. Étant donné le caractère sensible des données, ces corrections nécessitent une vérification d'un expert médical. En ce qui concerne la nature des erreurs, certaines étaient dues à l'absence de la forme particulière de transition que sont les options. D'autres corrections mineures ont été appliquées la plupart du temps : l'adaptation d'une forme ou d'une couleur, la suppression de racines inutiles, l'ajout de transitions, etc. Un exemple de correction plus complexe implique la présence d'un circuit. Elle était due à la signification du processus décrit : un examen médical doit être effectué plusieurs fois jusqu'à ce que le résultat soit satisfaisant. Jusqu'à présent, aucune formalisation satisfaisante n'a été trouvée pour exprimer ce type de connaissance en OWL.

Analyse de l'expressivité des arbres existants. Si KcatoS permet de formaliser la plupart des connaissances contenues dans les arbres, une expressivité accrue est nécessaire dans quelques cas. L'exemple précédent d'interdire les circuits révèle une limite du langage KcatoS. En outre, les connaissances décisionnelles peuvent également dépendre de divers facteurs tels que le temps, le calcul d'un score ou un ensemble complexe de critères. Pour prendre tous ces cas en compte, une extension du vocabulaire de KcatoS est nécessaire. Toutefois, le risque serait de rendre le système plus complexe et d'accroître les difficultés de compréhension pour les experts médicaux. Du point de vue de la formalisation, ces cas réfèrent souvent à des sujets complexes déjà abordés dans la littérature. Ainsi, traiter avec le temps dans le langage OWL est en partie rendu possible par l'utilisation de l'ontologie de concepts temporels OWL-Time [Pan et Hobbs, 2006]. En ce qui concerne les ensembles de critères, ils sont présents lorsque la décision dépend de nombreux facteurs. La décision finale repose généralement sur l'expérience des praticiens. Dans certains cas, il semble qu'une approche fondée sur la théorie des sous-ensembles flous serait utile. D'autres travaux du projet Kasimir ont déjà abordé ce sujet [d'Aquin *et al.*, 2006b].

5.6 Conclusion partielle

De nombreuses études récentes ont montré l'importance des GBPI en informatique médicale. Alors que plusieurs formalismes ont été proposés, aucun ne s'est imposé comme un standard universel. Dans ce chapitre, le *framework* KcatoS a été présenté. Fondé sur un langage de décision simple à appréhender, il propose une compatibilité avec des formats de GBPI établis tels que la syntaxe Arden et GLIF et avec les formats du web sémantique. Sa modularité permet une adaptation rapide aux besoins des utilisateurs.

Le *framework* propose également un éditeur collaboratif, intégrable à des projets de plus grande envergure tels qu'OncoLogiK. Couplé à d'autres outils du projet Kasimir tels que KOWL et EdHibou, il permet de poser les premières bases d'un système d'aide à la décision en oncologie exploitant les ressources du Web sémantique.

Déjà en production dans le cadre du projet OncoLogiK, l'expressivité du langage a été testée, ce qui a ouvert de nouvelles perspectives quant à son évolution.

Chapitre 6

Perspectives

Sommaire

À ce point d'avancement, plusieurs perspectives apparaissent tant pour OncoLogiK que pour le *framework* KCATOS. Le déploiement de l'application et sa mise en service ont fait émerger de nouveaux besoins. Le but de ce chapitre est d'évoquer l'ensemble de ces perspectives et de proposer de nouveaux objectifs réalisables pour le projet KASIMIR.

6.1 Évolutions d'OncoLogiK

OncoLogiK est désormais utilisé de manière quotidienne au sein du réseau Oncolor. Il remplit sa fonction de support pour les praticiens et est au centre du processus de mise à jour des GBPI. La souplesse des wikis sémantiques permet d'envisager de nombreuses évolutions et ouvre la porte à des solutions innovantes. Les perspectives étudiées pour OncoLogiK se concentrent sur deux niveaux. Le premier niveau traite de l'amélioration des requêtes bibliographiques. Le but est d'augmenter leur précision afin de fournir de meilleurs résultats aux utilisateurs. Le second niveau étudie la manière dont peut être améliorée la base de connaissances du wiki afin d'augmenter ses capacités.

6.1.1 Amélioration des requêtes bibliographiques

Dans le chapitre 3 est présentée la fonctionnalité d'OncoLogiK permettant de générer des requêtes bibliographiques ciblées vers les espaces de gestion des publications en ligne. En particulier, des requêtes sont dirigées vers PubMed afin d'interroger son moteur de recherche en fonction de mots-clés qualifiant le guide en cours de lecture. Ces mots-clés sont des termes issus du thésaurus MeSH et leur sélection a été réalisée par des experts médicaux. Dans la plupart des cas, l'annotation est constituée d'un ou deux termes se rapportant à la localisation des tumeurs décrites dans le guide. Les résultats obtenus fournissent en général un large éventail de réponses (en général, plusieurs dizaines de milliers sur PubMed) sans que ne soit garantie la complétude des informations fournies.

Le but de la première partie de cette section est d'évoquer des pistes afin d'améliorer la qualité des résultats de ces requêtes. Cependant, les optimisations proposées mènent à reconsidérer le processus d'indexation effectué par les experts, qui s'est avéré long et coûteux. La seconde partie propose donc des idées permettant de simplifier ce travail.

6.1.1.1 Comment améliorer la qualité des résultats des requêtes bibliographiques ?

Dans la littérature d'informatique médicale, l'interrogation des moteurs de recherche est un sujet récurrent. De nombreux travaux ont montré l'importance de leur utilisation. Ainsi, il est affirmé par Glasziou *et al.* [2008] que « l'utilisation d'un moteur de recherche est maintenant aussi essentielle que le stéthoscope » [47]. Plusieurs études se concentrent sur le moteur de PubMed, proposant des stratégies pour améliorer la qualité des résultats. Par exemple, l'apport du module *Clinical Query* qu'utilise OncoLogiK est montré par Lokker *et al.* [2011]. L'un des principaux enseignements est montré par la corrélation entre la construction de la requête et la qualité des réponses. Plusieurs pistes peuvent donc être évoquées pour améliorer la qualité des requêtes vers PubMed.

Intégrer les annotations en texte libre. D'après le tutoriel sur l'emploi des termes MeSH sur PubMed (NN/LM [2013b]), il est indiqué que la couverture des articles annotés avec ce

47. « *the use of search engines is now essential as the stethoscope* ».

thésaurus représente 90% des publications. Pour les autres publications, l'indexation repose sur le titre, le résumé et des mots-clés fournis par l'éditeur. Or, ces publications ne sont pas prises en compte lors de recherches se fondant exclusivement sur des termes MeSH. L'introduction de texte libre permettrait d'inclure ces publications dans les résultats des recherches.

Pour éviter d'augmenter le travail d'indexation des guides par les experts médicaux, une solution automatique peut être appliquée. En effet, le thésaurus MeSH propose des synonymes pour chacun des descripteurs, c'est-à-dire des expressions ayant un sens équivalent. Par exemple, le terme « Stomach Neoplasms » est utilisé pour indexer le guide « Estomac » sur OncoLogiK. La requête « "Stomach neoplasms"[Mesh] » vers PubMed proposée par OncoLogiK retourne 69 433 résultats. Or, le synonyme « gastric cancer » est donné dans le thésaurus. En l'intégrant à la requête qui devient « "Stomach neoplasms"[Mesh]OR"gastric cancer" », PubMed propose 76 468 résultats.

Une généralisation pourrait ainsi consister à intégrer tous les synonymes à OncoLogiK afin de permettre des requêtes plus complètes. Toutefois, il est intéressant de constater que de nombreux synonymes sont proposés pour chaque terme. Pour le terme évoqué précédemment, au total 21 synonymes sont proposés. Si la solution semble assez simple à mettre en place, intégrer tous ces synonymes conduirait toutefois à la création de requêtes illisibles, ce qui ne devrait pas correspondre aux attentes des utilisateurs. En effet, la requête est destinée à fournir une première approche pour l'utilisateur qui se charge ensuite de l'affiner en fonction de ses attentes. Une solution intermédiaire peut consister à intégrer dans le wiki uniquement les synonymes les plus utiles. Cela nécessiterait toutefois un imposant travail de tri dans le thésaurus MeSH.

Optimiser l'utilisation du moteur de PubMed. Une autre piste consiste à élargir les possibilités des requêtes en étendant le nombre de connecteurs disponibles. En effet, lorsqu'un guide est indexé par plusieurs termes, l'utilisateur dispose de la possibilité de choisir entre les connecteurs *AND* et *OR* pour les relier dans la requête. Or le moteur de recherche de PubMed offre d'autres possibilités, telles l'emploi du connecteur *NOT* qui permet d'exclure les publications indexées par certains termes. D'autre part, il est également possible d'associer à un terme l'annotation « [majr] » qui indique au moteur de ne retenir que les publications dont ce terme est le sujet principal.

Exploiter la richesse de MeSH. Le thésaurus MeSH est un outil complexe et complet, dont toutes les possibilités n'ont pas encore été exploitées par OncoLogiK. L'emploi des synonymes précédemment proposé en est un exemple. D'autres peuvent être évoqués. En effet, les termes de MeSH sont triés selon une hiérarchie. Or, cette particularité n'a pas encore été exploitée. Elle peut fournir des directions pour guider les utilisateurs vers des requêtes plus larges ou plus précises selon les besoins.

D'autre part, MeSH intégre une collection particulière de termes appelés *qualifiers*. Elle correspond à 83 termes consacrés à des sujets spécifiques dont l'association avec les descripteurs classiques permet d'aborder ces derniers sous un aspect particulier. Une liste commentée des *qualifiers* est disponible en ligne [NN/LM, 2013a]. Par exemple, l'emploi du *qualifier* « *surgery* » dans la requête « "Stomach neoplasms/surgery"[Mesh] » permet de préciser que les documents recherchés concernent les actes chirurgicaux appliqués dans les cas de cancer de l'estomac. L'emploi de ces termes permettrait ainsi de préciser les requêtes. Toutefois, les guides indexés dans OncoLogiK couvrent généralement plus d'un seul *qualifier*. Par exemple, le guide « Estomac » évoque le diagnostic, des classifications, des informations sur les traitements de radiothérapie, les actes chirurgicaux, etc. qui correspondent à autant de *qualifiers*. On peut remarquer que chacun

de ces aspects du guide correspond à une sous-partie du document. Il semblerait alors judicieux de relier l'annotation à cette sous-partie uniquement. Or ce processus n'est actuellement pas possible dans un wiki sémantique, puisque chaque annotation se rapporte au sujet de la page et non uniquement à une partie de celle-ci. Une solution serait alors d'intégrer au wiki sémantique un modèle de document. Cette proposition est développée dans la section 6.1.2.1.

Intégrer de nouvelles sources bibliographiques. Actuellement, les requêtes bibliographiques sont dirigées vers deux références, PubMed et Cismef. Celles-ci ont été choisies en accord avec les experts médicaux et semblent complémentaires : PubMed intègre de manière complète la littérature internationale tandis que Cismef propose des références francophones. Toutefois, de nombreuses autres sources sont disponibles, dont certaines peuvent sembler de prime abord utiles dans le contexte :

- *NHS Evidence* [NICE, 2013] archive et indexe de nombreuses publications sur les médicaments et les prescriptions ;
- *National Guideline Clearinghouse* [AHRQ, 2013] répertorie les GBPC écrits selon les principes de la médecine fondée sur les preuves ;
- L'organisme HON, responsable du label HONcode, propose un moteur de recherche ne prenant en compte que les sources qu'il certifie ;
- les généralistes *Google Scholar* [Google, 2013a] et *Intute* [Intute, 2013] indexent des publications scientifiques dans de nombreux domaines, dont la médecine.

L'ensemble de ces sites propose des moteurs de recherche interrogeables à travers des requêtes, de la même manière que Cismef et PubMed. Ils acceptent le texte libre, ce qui donne une raison supplémentaire de l'intégrer aux annotations d'OncoLogiK. Toutefois, avant d'être utilisées, toutes ces sources doivent être validées par les experts médicaux. Ensuite, un travail devra être réalisé afin d'optimiser les requêtes en fonction des moteurs, à la manière des propositions précédentes faites pour PubMed.

6.1.1.2 Comment faciliter le travail d'indexation des guides ?

La plupart des propositions précédentes ne peuvent être réalisées qu'en reprenant et approfondissant le travail d'annotation des guides effectués par les experts médicaux. Comme il est observé par Huang *et al.* [2011], l'annotation manuelle de publication par des termes MeSH est une tâche complexe, subjective et coûteuse en temps qui nécessite une compréhension de la publication concernée et une familiarité avec le thésaurus. Deux pistes sont évoquées pour simplifier le travail d'indexation. La première consiste à développer une interface pour simplifier le travail, la seconde propose d'y intégrer des algorithmes d'indexation automatique.

Améliorer l'ergonomie de la saisie des termes. La saisie des termes MeSH sur OncoLogiK repose sur l'emploi de modèles. Il n'existe cependant aucun guide pour aider les experts dans le choix de ces termes. Si on souhaite améliorer la précision des requêtes, il conviendra donc de guider les experts dans ces choix. Une interface ergonomique peut ainsi être développée pour simplifier ce travail. Un exemple peut être donné par le générateur de requêtes de PubMed montré dans la figure 6.1. Le but de cette interface est de préciser une requête en fonction d'un descripteur MeSH préalablement donné. Elle propose ainsi la liste des *qualifiers* disponibles et diverses options. Ce type d'interface peut être une inspiration pour la création d'une application pour l'aide à l'indexation.

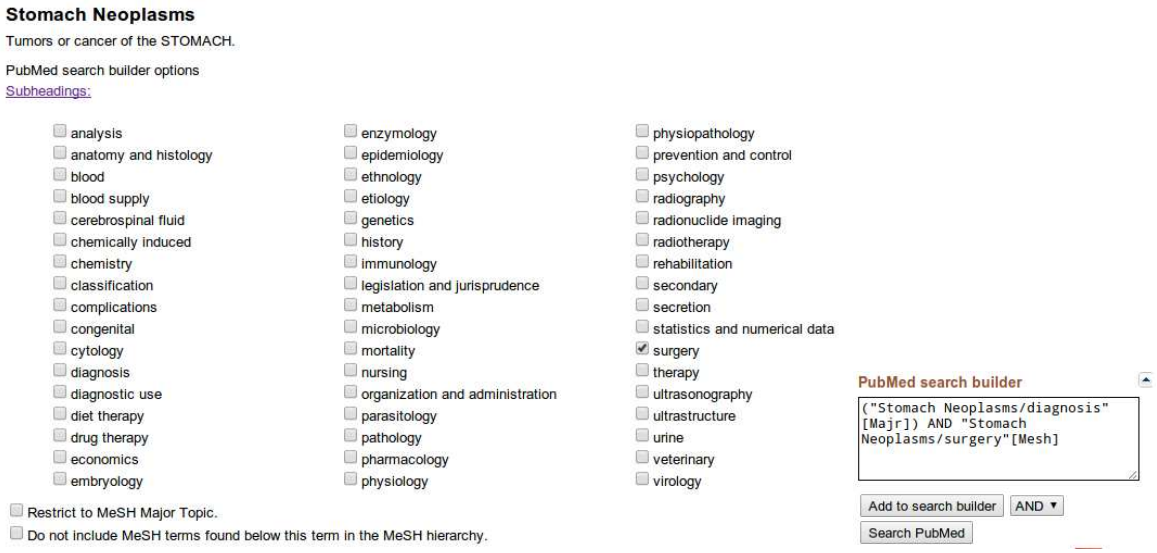


FIGURE 6.1 – Édition d'un requête sous PubMed à partir du descripteur *Stomach Neoplasms*. Source : NCBI [2013a].

Utiliser des méthodes automatiques d'indexation. Une autre piste pour faciliter le travail d'annotation peut être le recours aux méthodes automatiques d'indexation. Plusieurs outils et techniques sont décrits dans la littérature, comme par exemple par Huang *et al.* [2011] et par Pereira *et al.* [2008]. Ce dernier travail en particulier semble directement applicable aux guides contenus dans OncoLogiK puisqu'il décrit un outil nommé F-MTI qui propose des descripteurs MeSH à partir de résumés en langue française.

6.1.2 Amélioration de bases de connaissances de GBPI

Un an après la mise en place d'OncoLogiK, des premiers retours d'utilisateurs ont eu lieu. Ils ont mis en avant le besoin d'améliorer le confort d'utilisation de l'outil. Plusieurs propositions furent mineures et n'ont pas nécessité de grandes modifications. Elles furent principalement d'ordre ergonomique. D'autres ont mis en avant des problématiques impliquant un travail de recherche supplémentaire, nécessitant d'obtenir des connaissances complémentaires. Une première piste a déjà été évoquée dans le chapitre 3, en proposant l'intégration de données issues de bases externes telles que DBpedia. Deux autres pistes sont évoquées dans cette section. La première propose de prendre en compte les informations liées à la struture documentaire des GBPI, tandis que la seconde mène à s'interroger sur la part de contenu automatiquement formalisable.

6.1.2.1 Mieux orienter les spécialistes : vers un modèle de document ?

La communauté d'utilisateurs d'OncoLogiK est composée de spécialistes médicaux, de médecins généralistes et de patients. Pour chaque type d'utilisateur, la nature des informations recherchées varie. Ainsi, un chirurgien consultant le guide « Estomac » s'intéressera plus particulièrement à la section 12 nommée « Estomac », décrivant les différents types d'opérations chirurgicales applicables dans les cas de cancer de l'estomac. À l'heure actuelle, aucun outil ne peut guider l'utilisateur directement vers la spécialité qui l'intéresse. De la même façon, le module de requêtes

sémantiques ne pourrait pas répondre à une question triviale du type « Existe-t-il des traitements chirurgicaux pour le cancer de l'estomac ? ».

Une première réponse pourrait être de renforcer les annotations de chaque guide en ajoutant les *qualifiers* de MeSH, comme le propose la section 6.1.1.1. Toutefois, cette solution ne répond pas au problème de guidage de l'utilisateur. Pour cela il faudrait savoir où se trouvent précisément les informations intéressantes, à l'échelle de la page.

La manière dont sont annotés les guides dépend du paradigme des wikis sémantiques. Les guides édités dans un wiki forment des documents structurés, divisés en sections et contenant du texte, des illustrations graphiques, des arbres de décision, des références bibliographiques, etc. Lorsqu'ils sont intégrés à un wiki sémantique, les annotations qu'ils contiennent se rapportent au sujet général dans la page. Par exemple, l'annotation par le terme MeSH « Stomach Neoplasms » de la page « Estomac » permet la création d'un triplet du type :

```
(Oncologik:Estomac,
 Oncologik:aPourDescripteur,
 Mesh:Stomach Neoplasms)
```

Ainsi, le sujet du triplet est l'objet de la page. Or, dans le cas présent, l'annotation concerne une partie de la page et non la page entière.

Un modèle de document pour les wiki sémantiques. Pour parvenir à contourner cet obstacle, une solution consiste à intégrer aux wikis sémantiques un modèle de document. Le but de ce modèle serait de formaliser sa structure (c'est-à-dire son plan, etc.), autrement dit de créer le modèle des méta-données des pages de wiki. Ainsi, un paragraphe serait identifié comme une partie d'un document et il serait possible d'intégrer des annotations liées uniquement à ce paragraphe. On peut d'ailleurs noter que certains formats de GBPI présentés dans les chapitres précédents proposent ce type de modèle, par exemple GEM et Helen. Malheureusement, les modèles proposés sont décrits en XML et ne peuvent pas être utilisés par des moteurs d'inférences.

Dans un souci d'interopérabilité sémantique, il serait possible d'intégrer au modèle le vocabulaire du Dublin Core [DCMI, 2013]. Ce vocabulaire, objet d'une norme ISO et reconnu par le W3C, permet de décrire des ressources numériques et d'établir des relations entre celles-ci. Concrètement, il définit un certain nombre de propriétés et de classes utilisables pour décrire des pages Web telles que les propriétés « *title* » ou « *partOf* ». À titre d'exemple, le Dublin Core est utilisé pour décrire les descripteurs du thésaurus MeSH. Il est intéressant de noter que l'usage de ces propriétés, en particulier « *isReplacedBy* » ou « *hasVersion* », pourrait également contribuer à modéliser le processus de gestion des versions du wiki.

6.1.2.2 Quelle proportion du contenu des guides peut-on formaliser ?

Vers une formalisation exhaustive de GBPI ? Évoquée dans le chapitre 1, l'opposition entre les approches documentaires et les approches fondées sur les connaissances constitue l'un des points de débat pour la création des GBPI. L'une des originalités de la solution composée par OncoLogiK et KCATOS est de tenter de concilier ces approches, en exploitant les facilités d'édition des approches documentaires et les formalismes des approches fondées sur les connaissances. Ainsi, la première approche est utilisée pour les descriptions textuelles tandis que la seconde exploite les arbres de décision. La production automatique de bases de connaissances à partir de la seconde approche ouvre toutefois plus d'opportunités pour l'informatique médicale, permettant en particulier de proposer des bases pour les systèmes d'aide à la décision. Et dans ce cas, formaliser l'ensemble du GBPI permettrait d'obtenir des bases de connaissances plus complètes.

On peut donc se demander si cette approche est applicable à tout un GBPI, ce qui revient à chercher si, en l'état du processus d'édition, un GBPI peut être formalisé entièrement.

La majorité des personnes impliquées dans le processus d'édition des GBPI dans OncoLogiK sont des experts dont les compétences sont orientées vers la médecine. Or, la création de bases de connaissances complexes, comme celles qui peuvent contenir l'intégralité des GBPI, nécessite des compétences approfondies en ingénierie des connaissances. Ce profil de spécialiste n'est malheureusement pas inclus dans le processus d'édition. La politique du projet KASIMIR est de fournir aux utilisateurs les outils permettant la formalisation d'une manière la plus transparente et la moins invasive possible pour l'éditeur, de façon à ne pas provoquer un surplus de travail. Il faut donc chercher d'autres outils permettant de formaliser les contenus mais se gardant de trop alourdir la charge du travail d'édition.

Une approche possible est de proposer l'emploi d'un langage contrôlé dans les GBPI. Selon de Coi *et al.* [2009], les langages contrôlés sont des langages proposés pour simplifier les échanges entre les humains et les machines. Généralement constitués d'un fragment d'un langage naturel, ils sont composés d'un vocabulaire contrôlé, dans lequel les termes ne sont pas ambigus, et d'une grammaire formelle. Utilisé dans le cadre du moteur de wiki sémantique AceWiki présenté dans le chapitre 2, le langage ACE, fragment de l'anglais, est l'exemple d'un tel langage.

Si l'emploi d'une telle solution semble théoriquement possible, sa mise en application sera confrontée à des problèmes majeurs :

- À notre connaissance, il n'existe pas de langage contrôlé issu de la langue française permettant une couverture totale du vocabulaire utilisé dans les GBPI. Étant donné le jargon médical employé, il serait nécessaire de disposer d'un vocabulaire spécialisé.
- L'emploi d'un tel langage nécessiterait la création d'outils spécialisés afin d'assurer la bonne formation des phrases. Ce point peut être nuancé par les exemples d'interfaces proposés dans Acewiki.
- L'utilisation du langage nécessiterait une formation à large échelle pour tous les éditeurs des GBPC, ce qui induirait de nouveaux coûts.

Repérer les régularités de structure. Une perspective alternative et plus réaliste consiste à identifier d'autres structures formalisables automatiquement au moment de l'édition, à l'image des arbres de décision. De cette manière, on pourrait définir les besoins en outils logiciels permettant une édition à deux niveaux, un niveau identique ou proche de la structure actuellement employée et un niveau de représentation des connaissances. Pour identifier ces structures, des recherches doivent être menées dans les GBPI existants dans OncoLogiK. L'exemple d'une telle structure, celui des classifications TNM, est détaillée dans le paragraphe suivant.

Exemple du cas des classifications TNM. Les classifications TNM sont des systèmes de classement des cancers selon leur extension anatomique [UICC, 2013]. Elles permettent la description des tumeurs à travers 3 dimensions :

- la lettre « T », cotée généralement de T0 à T4, décrit la tumeur initiale selon le volume tumoral,
- la lettre « N », cotée généralement de N0 à N2, décrit le nombre de ganglions métastasés,
- la lettre « M » est généralement cotée M0 ou M1 selon selon la présence ou l'absence de métastases.

Les notations « Tx », « Nx » et « Mx » représentent l'absence d'information pour les dimensions respectives.

T - Tumeur primitive	
Tx	Tumeur primitive non évaluable
T0	Pas de signe de tumeur primitive
Tis	Carcinome *in situ*, maladie de Bowen, lésion intra-épithéliale squameuse de haut grade, néoplasie intra-épithéliale du canal anal (AIN II-III)
T1	Tumeur ≤ 2 cm dans sa plus grande dimension
T2	Tumeur >2 cm et ≤ 5 cm dans sa plus grande dimension
T3	Tumeur de plus de 5 cm dans sa plus grande dimension
T4	Tumeur envahissant un ou plusieurs organes de voisinage
N - Ganglion	
Nx	Envahissement ganglionnaire non évaluable
N0	Pas d'adénopathie régionale métastatique
N1	Adénopathies régionales métastatiques périrectales
N2	Adénopathies unilatérales iliaques internes et/ou inguinales
N3	Adénopathies métastatiques périrectales et inguinales et/ou iliaques internes bilatérales et/ou inguinales bilatérales
M - Métastase	
M0	Pas de métastase à distance
M1	Présence de métastases à distance

(a) Classification TNM extraite du GBPI « Anus ».

0	Tis	N0	M0
I	T1	N0	M0
II	T2-3	N0	M0
IIIA	T1-3	N1	M0
	T4	N0	M0
IIIB	T4	N1	M0
	tout T	N2, N3	M0
IV	tout T	tout N	M1

(b) Détail des stades de traitement.

FIGURE 6.2 – Classification TNM extraite du GBPI « Anus » d'OncoLogiK (6.2a) et détail des stades correspondants (6.2b).

Dans le cadre d'OncoLogiK, ces classifications sont régulièrement utilisées. Par exemple, la majorité des GBPI de la catégorie « Appareil digestif » en comporte une. La figure 6.2a en montre un exemple. Elles sont généralement accompagnées d'un tableau permettant de déterminer un « stade », pour lequel une recommandation est décrite à la section correspondante du document. Par exemple, le tableau correspondant à la classification TNM de la figure 6.2a est présenté dans la figure 6.2b.

Les régularités de ces classifications permettent d'envisager la création de bases de connaissances lors de leur édition. En effet, par exemple, à partir des lignes T1, N0 et M0 de la classification de la figure 6.2a, on peut aisément extraire les connaissances suivantes :

$$\texttt{SituationT1} \equiv \texttt{SituationX} \sqcap \exists \texttt{aPourTailleTumeur.} \leqslant_2$$

$$\texttt{SituationN0} \equiv \texttt{SituationX} \sqcap \ni \texttt{présenceAdénopathieRégMétastatique.faux}$$

$$\texttt{SituationM0} \equiv \texttt{SituationX} \sqcap \ni \texttt{présenceMétastaseDistant.faux}$$

Puis, en suivant la logique de construction des classifications TNM, on peut écrire :

$$\texttt{SituationT1M0N0} \equiv \texttt{SituationT1} \sqcap \texttt{SituationM0} \sqcap \texttt{SituationN0}$$

Enfin, en exploitant la ligne I du tableau de la figure 6.2b, on peut écrire :

$$\texttt{SituationT1M0N0} \sqsubseteq \exists \texttt{aPourStade.StadeI}$$

Cet exemple élémentaire montre que des connaissances peuvent être extraites simplement des classifications. Il convient maintenant de définir un algorithme général pour leur extraction, et de proposer des outils d'édition adaptés au contexte des wikis sémantiques pour la construction semi-automatique de ces bases.

6.2 Évolutions de KCATOS

Dans le chapitre précédent, la confrontation des fonctionnalités du *framework* KCATOS à celles des formats existant laisse apparaître de nombreuses possibilités d'évolution. Cette section présente des évolutions souhaitables, d'abord au niveau de l'expressivité du langage, ensuite au niveau des environnements d'exécution.

6.2.1 Expressivité du langage

Les possibilités d'expression des connaissances dans KCATOS sont fortement liées à l'emploi des technologies du Web sémantique. Ainsi, certaines limites d'OWL constituent également des limites pour les arbres de décision. L'interdiction des circuits dans ces arbres en est un exemple. La question des possibilités est évoquée dans la section 6.2.1.2. Cependant, KCATOS est loin d'utiliser toutes les ressources de ce langage de représentation, ce qui ouvre la voie à de nombreuses améliorations. Ainsi, la section suivante présente les possibilités d'intégrer des représentations temporelles.

6.2.1.1 Comment intégrer le temps ?

Comme le souligne les différents états de l'art des GBPI évoqués dans le chapitre 5, la représentation des contraintes temporelles dans les GBPI est une problématique centrale. D'ailleurs,

	Asbru	EON	GLIF	PRO*forma*	KcatoS
Basic control-flow	5	5	5	5	5
Advanced branching and synchronization	2	2	2	3	3
Structural patterns	1	2	2	1	1
Multiple instances patterns	0	0	0	0	0
State-based patterns	2	0	1	2	1
Cancellation patterns	2	1	1	2	0
New patterns	8	1	6	9	1
	20	11	17	22	11

TABLE 6.1 – Support des modèles de *Control-flow* pour Asbru, EON, GLIF, PRO*forma* et KcatoS. Source : Mulyar *et al.* [2007].

plusieurs formats proposent des mécanismes de gestion, comme par exemple SAGE et PRO*forma*. Asbru propose même la gestion d'intervalles flous afin d'offrir plus de flexibilité.

Dans le cadre des GBPI du projet KASIMIR, l'apport des contraintes temporelles apparaît en deux points :

- lors de l'utilisation de plusieurs recommandations pour une même situation, si l'ordre de ces recommandations a une importance,
- pour pouvoir représenter des délais sur les liens entre figures.

La formalisation de ces connaissances implique de pouvoir représenter des intervalles de temps, des points temporels relatifs, des mécanismes d'ordonnancement et de disposer des types de données relatifs au temps.

D'une part, l'intégration des types de données relatifs aux dates dans OWL 2 [Motik *et al.*, 2009] permet de répondre à l'un des points. D'autre part, certaines ontologies proposent des représentations de concepts temporels. Un exemple est présenté par Batsakis et Petrakis [2011], qui proposent l'ontologie S-OWL destinée à la représentation de données spatio-temporelles. Utilisant OWL 2 et des règles SWRL, elle semble proposer un modèle de données intégrant les besoins de représentation évoqués plus tôt. Ce modèle utilise des raisonnements décidables avec une expressivité correspondant à la LD $\mathcal{SHRIF}(D)$.

6.2.1.2 Des limites dans les parcours : intégrer les possibilités des *workflows* ?

Dans certains arbres de décision sont utilisés des circuits, c'est-à-dire la représentation dans les arbres de parcours appelés à se répéter tant qu'une condition de fin n'est pas remplie. Or, comme il est dit dans le chapitre précédent, le langage KcatoS ne parvient pas à modéliser ces parcours, ce qui constitue donc une de ses limites. Il est donc légitime de se poser la question sur d'autres limites de parcours pouvant apparaître à l'avenir.

Pour imaginer l'ensemble des parcours possibles dans ce type de structure, il est intéressant de se pencher sur les recherches effectuées dans le domaine des *workflows*. En particulier, Russell *et al.* [2006] proposent un ensemble de modèles dits de *Control-flow* destinés à identifier le maximum de structures utiles pour la modélisation des flux de données. En tout, 43 modèles sont proposés, groupés en 7 catégories :

- la catégorie « *Basic control-flow* » propose des modèles élémentaires, tels que les choix exclusifs ou le déroulement d'actions en séquences ;
- la catégorie « *Advanced branching and synchronization* » décrit les comportements possibles entre deux actions, typiquement les liens ;

- la catégorie « *Structural patterns* » étudie des possibilités générales du formalisme de modélisation ;
- la catégorie « *Multiple instance patterns* » décrit le fonctionnement de plusieurs instances d'exécution pour un même processus ;
- la catégorie « *State-based patterns* » caractérise des scénarios de processus dans lesquels des actions sont exécutées en fonction de l'état de l'instance ;
- la catégorie « *Cancellation patterns* » décrit des scénarios où une réinitialisation de l'instance est nécessaire ;
- la catégorie « *New patterns* » forme un ensemble hétéroclite de modèles complémentaires proposés par les auteurs.

Mulyar *et al.* [2007] proposent de se servir de ces modèles afin d'évaluer les formats de GBPI. Le but est alors de déterminer quels sont les modèles qu'Asbru, EON, GLIF et PRO*forma* sont capables d'utiliser lors de l'exécution des GBPI. Les résultats de cette étude sont résumés dans la table 6.1.

Si on applique cette évaluation à KcatoS, les résultats obtenus montrent que, sur 43 modèles proposés, il n'en implémente que 11. Le langage KcatoS se place à égalité avec le langage EON, mais largement en-dessous des langages Asbru, GLIF et PRO*forma*. Ce résultat peut s'expliquer par une différence importante lors de la conception. Alors que le langage KcatoS s'est assuré tout au cours de son développement d'être utilisable dans le contexte du Web sémantique, les autres formats ont été conçus en s'inspirant des systèmes de *workflows*, proposant des langages de représentation dédiés. Ces langages ont donc été adaptés pour offrir la souplesse nécessaire, mais y ont perdu leur généricité.

S'il n'est pas question de transformer le langage KcatoS en langage de *workflow*, il est tout de même intéressant d'étudier ces modèles, dont certains offrent des perspectives pour l'expression des connaissances des GBPI. On pensera notamment aux systèmes de *triggers* [48] qui permettent de suivre l'évolution d'une donnée particulière lors de l'exécution du GBPI et de déclencher une action si la valeur atteint un seuil prédéterminé.

6.2.2 Exécution des guides

Même si elle ne représente pas le point central de ce travail, l'exécution des guides est le but ultime de leur création. Il est donc nécessaire d'y accorder une grande attention. Actuellement, cette étape est gérée par le *framework* EdHibou. Celui-ci peut être complété par d'autres outils amenant une valeur ajoutée au projet. Ainsi, la section 6.2.2.1 propose de créer un lien avec les systèmes de dossier patient et la section 6.2.2.2 évoque la création d'un environnement de test pour les GBPI.

6.2.2.1 Peut-on intégrer le dossier patient ?

Parmi les formats présentés dans le chapitre 4, la plupart proposent des interactions avec un système de dossier patient. Son intégration permet d'envisager de proposer automatiquement des recommandations personnalisées en fonction des enregistrements de données médicales. Isern et Moreno [2008] citent l'inclusion du dossier patient dans les systèmes de GBPI comme l'un des points encore épineux, notamment à cause des problèmes d'interopérabilité. Pour que les systèmes puissent communiquer, il est indispensable d'employer un langage commun. Peleg *et al.* [2008], qui étudient le lien du format GLIF avec un système de dossier patient, affirment que

48. en français, déclencheurs.

cette intégration nécessite deux étapes : (1) lier les données à travers des ontologies d'application et (2) proposer des requêtes automatiques vers la base de données du dossier patient. Pour de nombreux formats de GBPI, le problème a été éludé en proposant des modèles simplifiés du dossier patient, permettant ainsi simplement de récolter les données. Ainsi, intégrer le dossier patient aux formats de GBPI reste un problème ouvert pour la recherche et l'ingénierie.

Plusieurs systèmes d'information existent pour le gestion du dossier patient. Des solutions *Open Source* sont présentées par Leong *et al.* [2007]. Développée dans le cadre d'un projet européen, *openEhr* [Beale et Heard, 2008] semble être une des plus complètes et des mieux documentées. Les enregistrements médicaux y sont stockés sous forme d'*archetypes*. Ces archétypes sont des structures destinées à définir et modéliser les informations liées à un enregistrement clinique. Elles sont prévues pour contenir des informations ontologiques destinées à favoriser l'interopérabilité avec d'autres systèmes. Ainsi, de nombreux archétypes contiennent des références à l'ontologie SNOMED CT (présentée dans le chapitre 1).

En imaginant l'emploi d'*openEhr* dans le cadre de KCATOS, et plus généralement du projet KASIMIR, il est donc nécessaire de proposer des annotations SNOMED CT pour lier les deux systèmes. Toutefois, le contexte du projet favorise l'adoption du thésaurus MeSH pour l'annotation des arbres, puisque ce thésaurus est déjà utilisé dans OncoLogiK. Pour ne pas en venir à un travail de double annotation des arbres de décision, il faut étudier les possibilités d'alignement entre MeSH et SNOMED CT. Jacobs *et al.* [2006] ont montré qu'un alignement automatique reposant sur le metathésaurus UMLS est possible pour plus de 90% des termes.

Au final, pour envisager un lien entre les arbres de décision de KCATOS et le système de dossier patient *openEhr*, plusieurs étapes sont nécessaires :

- intégrer des annotations sémantiques MeSH aux arbres de décision ;
- améliorer l'alignement entre MeSH et SNOMED CT ;
- proposer des requêtes constituées à partir des termes SNOMED vers la base de données d'*openEhr*.

6.2.2.2 Comment intégrer un environnement de test ?

Dans leur état de l'art des formats des GBPI, de Clercq *et al.* [2004] soulignent l'importance d'un environnement de test des GBPI. Son but est de désambiguïser et de valider aussi bien syntaxiquement, sémantiquement que médicalement, les GBPI. Ainsi, plusieurs niveaux de test sont nécessaires. Le premier niveau est syntaxique, de manière à garantir la bonne composition des arbres en fonction de la grammaire du langage. Le deuxième niveau est sémantique. Son but est d'abord de détecter des contradictions logiques ou d'éventuelles erreurs de procédure. Une erreur de procédure peut être, par exemple, de proposer un produit sans vérifier une allergie chez le patient ou de proposer des doses supérieures aux seuils de tolérance. Ces tests impliquent donc l'utilisation de connaissances médicales complémentaires à celles formalisées dans l'arbre de décision. Le dernier niveau consiste à simuler l'exécution de l'arbre dans des cas réels ou représentatifs, afin de faire vérifier le bien-fondé des recommandations proposées par des experts.

Dans le cadre de KCATOS, seule une partie de ces tests est prise en charge. En effet, des algorithmes sont proposés pour vérifier la syntaxe finale des arbres. De plus, des tests de cohérence sont réalisés sur les bases de connaissances extraites. Toutefois, seul le niveau syntaxique est vérifié. Du fait de l'absence d'annotation sémantique pour chaque figure composant le langage, il est actuellement impossible de détecter une répétition ou une contradiction médicale.

Deux types d'erreurs au niveau sémantique sont automatiquement identifiables. Le premier type contient les erreurs dont l'identification est possible en ne prenant en compte que la base de connaissances seule. Des exemples d'erreurs de ce type peuvent être la répétition de con-

naissances ou une contradiction apparaissant dans l'arbre. Leur identification dépendra alors principalement de la qualité des annotations médicales qui leur seront associées. Le second type d'erreur est celui pour lesquelles la détection nécessite une base de connaissances complémentaire. Par exemple, savoir que deux traitements ne sont pas compatibles et que leur administration commune peut faire courir des risques à la santé du patient. Pour ces cas, une base de connaissances complémentaire doit être utilisée, ce qui correspond à un lourd travail d'ingénierie médicale. Sa création pourrait reposer des techniques d'acquisition semi-automatiques, telles que le propose par exemple Bendaoud *et al.* [2008]. Un lien avec une telle base constituerait toutefois un apport scientifique conséquent et une garantie supplémentaire de qualité des GBPI.

Une autre base de test est généralement utilisée : son but est de confronter les GBPI à des cas médicaux afin de faire vérifier les recommandations déduites par des experts médicaux. Le but est alors de proposer une base de test comportant des cas typiques, mais également des cas pour lesquels les données correspondent à des valeurs extrêmes ou incomplètes. Dans ce sens, un lien avec le système de dossier patient aurait un double avantage : celui de proposer une interface pour la génération de cette base de cas et celui de pouvoir directement réaliser des tests comparables au déroulement dans le système d'exécution.

6.3 Perspectives pour le projet KASIMIR

Les perspectives présentées jusqu'à présent ont des répercutions localisées soit à OncoLogiK, soit au *framework* KCATOS. Cependant, l'apport de ces outils peut amener à repenser l'ensemble du projet KASIMIR. Il faut d'abord s'assurer que les bases de connaissances nouvellement créées peuvent fonctionner conjointement. C'est le but de la première section. La deuxième section propose pour sa part une nouvelle architecture logicielle et un nouveau modèle de données pour le projet KASIMIR, prenant en compte ses évolutions récentes et les perspectives proposées.

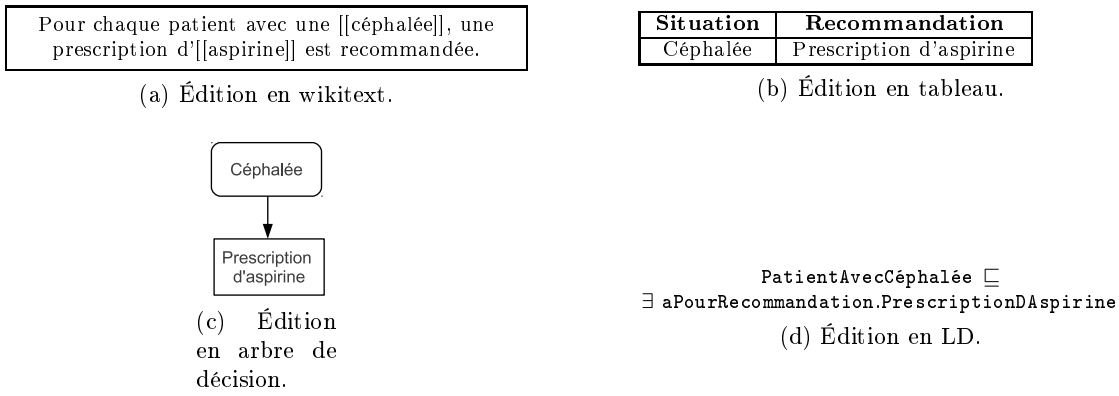
6.3.1 Unification des niveaux de connaissances

Les niveaux de représentation. À ce moment de l'évolution du projet KASIMIR, une particularité d'OncoLogiK est de proposer deux niveaux de représentation :

- Le premier niveau est celui de l'annotation des référentiels, très simple et peu contraint. Ce niveau exploite une terminologie externe (MeSH) et peut donc être relié à des services Web (comme dans le cas des requêtes vers CiSMEF et PubMed).
- Le second niveau est celui des arbres dont la structure suit des règles établies par un algorithme. Il ne comporte quasiment aucun lien vers l'extérieur (exception faite des liens explicites vers le wiki).

Les bases de connaissances sont reliées par une propriété éditée dans OncoLogiK qui indique l'utilisation d'un arbre de décision dans un GBPI ainsi que des propriétés indiquant un lien d'un arbre de décision vers un GBPI. Les niveaux d'édition sont complémentaires mais différents par leur expressivité (plus limitée dans le wiki que dans les arbres) et le type de connaissances formalisées. De plus, d'autres connaissances sont également éditées sans être formalisées, comme le montre le cas des classifications médicales ou, tout simplement, des textes et des schémas.

La faiblesse des liens entre les niveaux de représentation des connaissances constitue un frein au développement de raisonnements complexes en n'intégrant pas toutes les connaissances du projet KASIMIR. Par exemple, le système de recherche du wiki ne prend pas en compte les connaissances éditées dans les arbres, ce qui conduit à une perte de résultat lors de recherches d'informations.



(a) Édition en wikitext.

(b) Édition en tableau.

(c) Édition en arbre de décision.

(d) Édition en LD.

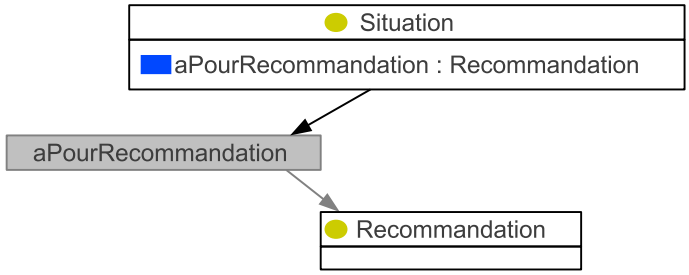
FIGURE 6.3 – Édition d'une connaissance à différents niveaux du continuum de formalisation des connaissances.

Le continuum de formalisation des connaissances. L'existence de ces différents niveaux de représentation peut être vue comme une particularité des wikis sémantiques. Ce phénomène a été décrit par Baumeister *et al.* [2011a] sous le nom de continuum de formalisation des connaissances (CFC). En effet, dans un domaine particulier tel que celui des GBPI, les mêmes connaissances peuvent être éditées par différents types de représentation. Ces représentations peuvent être proches, comme par exemple le texte libre et le texte dans un vocabulaire contrôlé, ou très éloignées, comme le texte et les règle logiques. Ainsi, dans le cadre d'OncoLogiK, plusieurs représentations sont présentes : texte libre, modèles de wikitext représentant les propriétés, tableaux contenant les classifications, arbres de décision en langage KCATOS, etc. Entre chaque mode de représentation, le degré de formalisation change graduellement mais la connaissance formalisée reste la même, celle du domaine. Le CFC peut ainsi être défini comme un concept mental représentant de cette évolution du degré de formalisation. La figure 6.3 illustre ce phénomène avec l'édition à plusieurs niveaux de représentation des connaissances déjà évoquées selon lesquelles une prescription d'aspirine est recommandée dans les situations de céphalée. Cette assertion peut être écrite en wikitext (figure 6.3a), sous forme de tableau (figure 6.3b), d'arbre de décision (figure 6.3c) ou en logique de descriptions (figure 6.3d). Chacune de ces formes représente un niveau de formalisation différent.

Ainsi, le CFC n'est ni un modèle physique, ni une méthodologie de développement de bases de connaissances mais plutôt une métaphore du processus de développement des connaissances. Le but est d'aider les experts à voir les données comme des connaissances qui peuvent être graduellement formalisées selon les besoins. Dans la suite de leur étude, Baumeister *et al.* [2011a] montrent que les wikis sémantiques sont les outils idéaux pour stocker les différentes représentations de ce continuum, étant donné leur propension à accepter plusieurs de ces niveaux.

Du point de vue de la gestion des connaissances, exploiter simultanément les connaissances issues des différents niveaux du CFC serait un apport important, permettant d'étendre les bases. Pour que ce soit possible, il est nécessaire d'identifier une structure assez générale pour identifier ces connaissances aux différents niveaux et de proposer une traduction dans un langage automatiquement exploitable. Il semble que les *Content Ontology Design Patterns* (CP) puissent être une telle structure.

Ontology Design Patterns. Présentés par Gangemi et Presutti [2009], les CP sont un type particulier d'*Ontology Design Patterns* (ODP). Un ODP est une solution de modélisation des-



Nom Situation

Intention Le but de ce modèle est de proposer une manière de relier une situation médicale aux recommandations qui lui correspondent.

Domaine Guides de bonnes pratiques informatisés

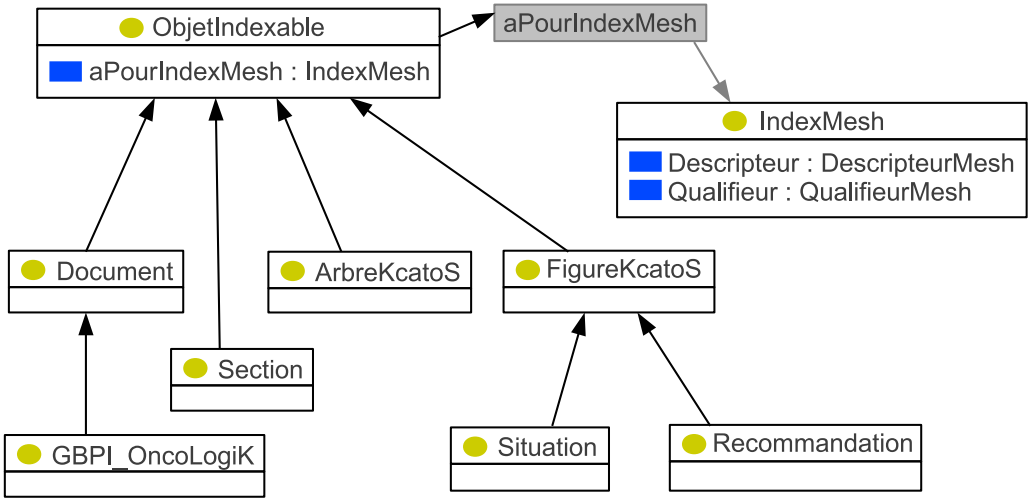
Description Pour toute situation de la classe `Situation`, il existe une recommandation de la classe `Recommandation` qui est associée à `aPourRecommandation`. Cette propriété `aPourRecommandation` lie une situation à une recommandation : elle a la classe `Situation` pour domaine et la classe `Recommandation` pour co-domaine.

Référence `http://labotalc.loria.fr/~kasimir/kcatos/2012/Situation.owl`

FIGURE 6.4 – Description du CP représentant le lien entre situation et recommandation dans l'algorithme d'export de KCATOS.

tinée à résoudre un problème récurrent de conception d'ontologie et peut ainsi être vu comme un ensemble de recommandations et de bonnes pratiques pour la conception d'ontologies. Ce concept très générique regroupe plusieurs familles de solutions destinées à simplifier le travail des ingénieurs des connaissances et à optimiser les bases, tant au niveau de la lisibilité que du raisonnement. Parmi ces familles, les *Logical OP* ont pour but de proposer des compositions de structures logiques pour résoudre un problème d'expressivité, indépendamment du domaine d'application. Un exemple de *Logical OP*, documenté dans un wiki spécialisé [OntologyDesignPatterns.org, 2010], propose une modélisation des relations n-aires. Un autre exemple de famille est la famille des *Architectural OP* dont plusieurs sont proposés par Ben Abbès *et al.* [2012] et qui ont pour but de décrire la forme générale d'une ontologie. Pour leur part, les CP proposent des modèles pour résoudre des problèmes de représentation en prenant en compte les classes et les propriétés du domaine d'intérêt. Il sont indépendants du langage de représentation, bien que leur utilisation dans le contexte du Web sémantique impose qu'il soit possible de les représenter en OWL. Leur but est ainsi de représenter un cas d'utilisation générique.

Les CP peuvent être appliqués simplement aux représentations utilisées par le langage KCATOS. En effet, la figure 6.4 propose un exemple de CP décrivant la relation entre les classes `Recommandation` et `Situation` utilisées dans l'algorithme d'export. La manière dont le CP est présenté (Schéma UML, nom, intention, etc.) est directement inspirée des propositions de Gangemi et Presutti [2009]. On peut noter que de la même manière, toutes les étapes de l'algorithme de traduction et les extensions existantes pourraient être modélisées par des CP. La création des bases de connaissances à partir des arbres reviendrait ainsi à un assemblage de plusieurs CP,



Nom Indexation par MeSH

Intention Le but de ce modèle est de proposer une manière d'unifier les niveaux de connaissances d'OncoLogiK.

Domaine Guides de bonnes pratiques informatisés

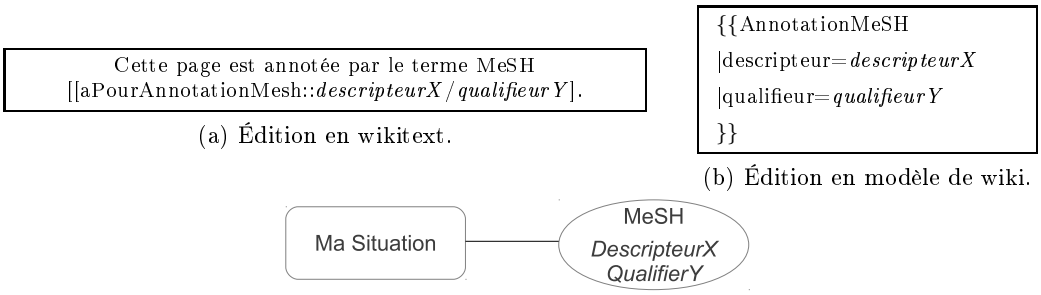
Description Les documents, ainsi que leur composants (sections, arbres de décision, figures des arbres) sont représentés par la classe `ObjetIndexable` dont ils sont des sous-classes. Les membres de cette classe sont des objets qui peuvent être annotés par le thésaurus MeSH. Une annotation par MeSH est composée d'un descripteur et d'un qualifieur.

FIGURE 6.5 – Description du CP « Indexation par MeSH » représentant la méthode d'unification des niveaux de connaissances dans OncoLogiK.

comme le préconise Gangemi [2005]. Ainsi, les bases gagneraient en lisibilité et simplifieraient également leurs liens avec des bases externes.

Application à OncoLogiK. Dans le cadre du CFC que forme OncoLogiK, la généricité des CP peut être utilisée pour définir une structure permettant d'unifier les différents niveaux de représentation. Comme évoqué précédemment, les CP sont indépendants du langage de représentation. Un même CP peut donc avoir plusieurs représentations en fonction du mode et du langage d'édition utilisés. Toutefois, à cette représentation du CP correspondra une représentation OWL unique, quelles que soient les modalités d'édition.

Pour créer un CP dans OncoLogiK, il convient d'identifier une régularité dans les différents niveaux de représentation, c'est-à-dire un type de connaissance utile et représentable dans les objets du CFC. Plusieurs des perspectives de ce chapitre évoquent le besoin d'augmenter et de préciser les annotations MeSH, tant au niveau des pages que des arbres de décision. Pour lier ces deux niveaux, on peut imaginer un CP représentant les annotations sémantiques des différents objets utilisés. L'exemple d'un tel CP est donné dans la figure 6.5. Ce CP pourrait être décliné selon plusieurs représentations, comme le montre la figure 6.6. Ces représentations plus accessibles permettraient une annotation plus simple, favorisant leur utilisation par les rédacteurs.



(a) Édition en wikitext.

(b) Édition en modèle de wiki.

(c) Édition dans le langage KCATOS.

FIGURE 6.6 – Propositions de déclinaison du CP « Indexation par MeSH » sous forme de wikitext (6.6a), de modèle de wiki (6.6b) et de figure du langage KCATOS (6.6c).

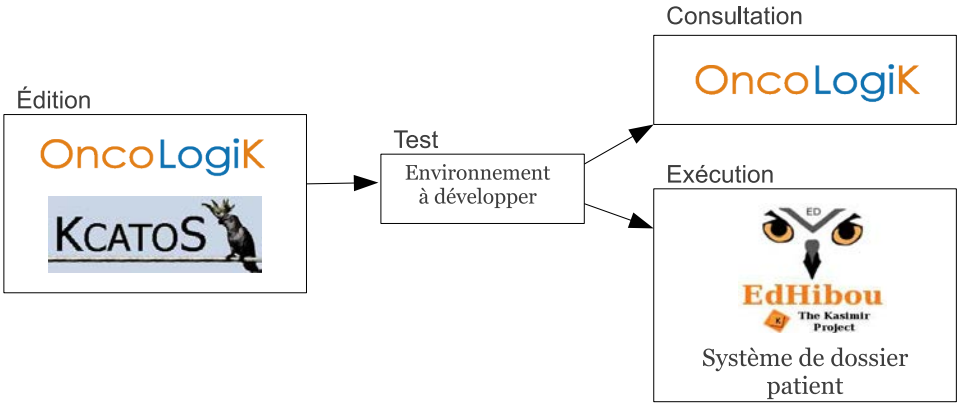


FIGURE 6.7 – Proposition d'architecture logicielle pour le projet KASIMIR.

6.3.2 Vers une nouvelle architecture pour le projet KASIMIR

En prenant en compte l'existant et les perspectives évoquées dans ce chapitre, il est possible de déterminer une nouvelle architecture pour le projet KASIMIR. Le but de cette architecture est de faire de KASIMIR un système d'aide à la décision complet. Ainsi, sont intégrés les différents composants nécessaires selon trois points de vue déterminés par de Clercq *et al.* [2004] : l'acquisition, le test et l'exécution/consultation.

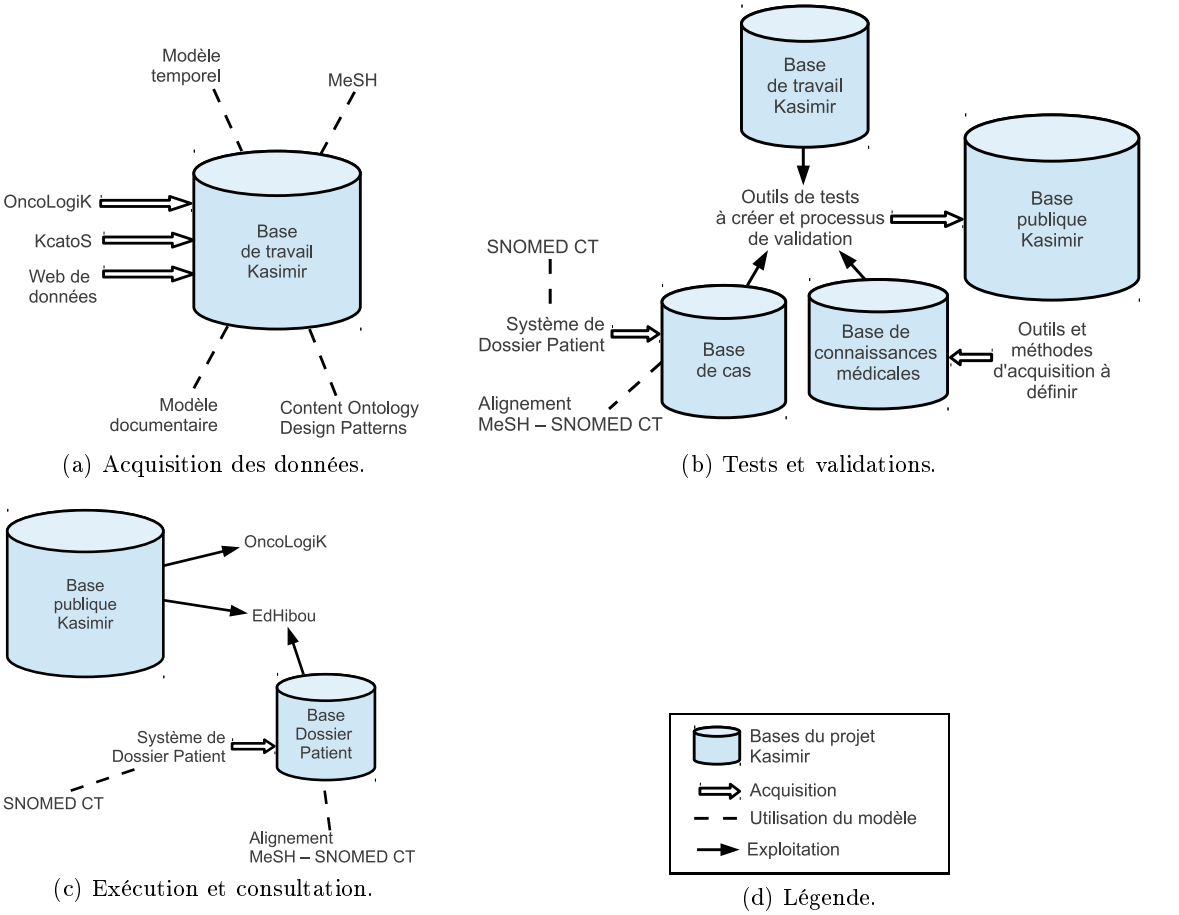
Deux types de modèles sont proposés. Le premier présente les environnements logiciels mis en œuvre et le second présente la manière dont les données y sont manipulées.

6.3.2.1 Architecture logicielle

L'acquisition de données repose sur les deux systèmes mis en place au cours de ce travail : OncoLogiK et KCATOS.

En ce qui concerne l'environnement de test, le développement d'un logiciel spécifique est nécessaire, et peut suivre les pistes présentées dans la section 6.2.2.2.

Pour consulter les GBPI, OncoLogiK fournit une interface de navigation confortable, à laquelle s'ajoutent des services Web tels que les requêtes vers les sites de bibliographie et les apports des wikis sémantiques. L'exécution des GBPI édités avec KCATOS pourra se faire via EDHIBOU,



(a) Acquisition des données.

(b) Tests et validations.

(c) Exécution et consultation.

(d) Légende.

FIGURE 6.8 – Propositions de modèles présentant l'acquisition (6.8a), le test (6.8b) et l'exploitation des données (6.8c) dans le projet KASIMIR.

en prenant en compte un système de dossier patient, comme par exemple openEHR.

Un schéma présentant cette architecture est proposé dans la figure 6.7.

6.3.2.2 Modèles de données

Acquisition. La construction de la base de travail du projet KASIMIR repose sur l'acquisition de données par KCATOS et OncoLogiK. Moyennant des validations postérieures, l'intégration de données du Web de données est envisageable. Plusieurs modèles gouvernent la structure des connaissances stockées :

- un modèle gérant les aspects temporels, tel que proposé dans la section 6.2.1.1,
- un modèle de document décrivant les méta-données des GBPI édités dans OncoLogiK,
- des CP assurant le lien entre les données issues des arbres et celles issues du wiki.

Enfin, les ressources sont indexées par le thésaurus MeSH.

Test. L'environnement de test a pour but d'intégrer le processus de validation des connaissances médicales et d'y ajouter un processus de validation automatique. S'ils restent à développer, plusieurs de ses composants peuvent déjà être décrits. Deux bases peuvent servir de support aux tests :

- une base de connaissances médicales générales : ses contenus sont encore à définir, mais il serait intéressant d'envisager sa création grâce à des méthodes semi-automatiques fondées sur le Web de données ;
- une base de cas, créée à partir d'un système de dossier patient.

Au final, une base publique est créée et maintenue à jour, offrant une fiabilité accrue.

Exécution/Consultation. La consultation des guides sur OncoLogiK permet d'exploiter les données pour proposer des services Web à l'utilisateur. La principale innovation proposée reposera sur l'utilisation par EdHibou de données issues du dossier patient. Pour ce faire, si openEHR est utilisé, un alignement entre MeSH et SNOMED CT devra être intégré.

6.4 Conclusion partielle

Dans ce chapitre, des perspectives sont présentées pour favoriser l'acceptation et l'évolution des outils proposés précédemment. Pour OncoLogiK, plusieurs évolutions possibles sont citées :

- l'amélioration de la précision des requêtes vers les sites de bibliographie, notamment en optimisant l'exploitation du thésaurus MeSH et en simplifiant son utilisation,
- la création d'un modèle de document pour les wikis sémantiques afin de prendre en compte les informations liées à la nature documentaire des GBPI,
- la recherche d'autres structures formalisables automatiquement, à l'image des classifications TNM.

Au niveau du *framework* KcatoS, les évolutions concernent les limites d'expressivité des arbres de décision et l'exécution des guides. Pour améliorer l'expressivité, d'une part, l'observation des apports d'autres solutions de GBPI nous conduit à intégrer des représentations temporelles, ce qui est possible dans le cadre d'OWL 2. D'autre part, de nouveaux parcours peuvent être définis, notamment en s'inspirant des modèles créés à destination des *workflows*.

Les interactions entre ces deux systèmes sont pour le moment relativement faibles. Pour y remédier, il est montré que l'utilisation d'ODP pourrait favoriser ces interactions.

Au final, une nouvelle structure pour le projet Kasimir est proposée, prenant en compte les outils existants et les perspectives précédemment évoquées. Une architecture logicielle est détaillée, ainsi que le modèle de données sous-jacent.

Conclusion générale

Les connaissances décisionnelles sont un type particulier de connaissances dont le but est de décrire des processus de prise de décision. En cancérologie, ces connaissances sont généralement regroupées dans les recommandations de pratique clinique (RPC). Ces documents pédagogiques proposent des solutions thérapeutiques en associant des recommandations cliniques à une situation médicale donnée. Leur publication est assurée par des organismes médicaux certifiés, suite à un processus d'édition complexe faisant intervenir des experts issus de différentes spécialités de la médecine et des infographistes. Ce processus est le reflet d'un travail collaboratif au cours duquel des étapes de validation et de mise à jour sont nécessaires. Ainsi, les documents finaux comportent des connaissances fiables représentant l'état de l'art au moment de leur publication.

La qualité des documents et l'évolution des technologies en informatique médicale a ouvert une nouvelle perspective pour les RPC, consistant à créer des guides de bonnes pratiques informatisés (GBPI), plus simples à diffuser. De plus, la création de versions électroniques a conduit à la volonté de formaliser l'ensemble des connaissances contenues dans le guide, de manière à pouvoir alimenter des bases de connaissances utilisables dans le cadre des systèmes d'aide à la décision. Ainsi, la création de ces contenus formels peut être vue commme une problématique d'acquisition des connaissances. Elle nécessite deux niveaux d'expertise : le premier niveau est médical et le second concerne l'ingénierie des connaissances.

Dans ce contexte, le but du travail présenté dans cette thèse était de proposer des méthodes et des outils permettant de factoriser l'édition des GBPI et leur formalisation. En d'autres termes, il convenait de proposer des méthodes permettant une acquisition des connaissances automatique lors de l'édition des GBPI par les experts médicaux. Pour répondre à cette problématique, trois apports sont proposés.

Le premier apport est l'intégration des technologies du Web social et sémantique dans le processus d'édition des GBPI. L'idée du Web social et sémantique est d'exploiter l'intelligence collective du Web social avec les technologies du Web sémantique. Ainsi, le rôle du Web sémantique est de fournir les techniques pour gérer la communication des connaissances tandis que le Web social favorise l'agrégation des contributions des utilisateurs. La création du wiki sémantique OncoLogiK a permis de mettre en œuvre cette proposition. Ainsi, un retour d'expérience et des méthodes sont présentés pour la migration d'une solution Web statique vers une solution du Web social et sémantique. De plus, les apports de cette migration sont présentés : la simplification et l'optimisation du processus, ainsi que l'ajout de services sémantiques.

Le second apport consiste à proposer une solution pour exploiter les connaissances décisionnelles présentes dans les GBPI. Le plus souvent, ces connaissances sont présentées sous forme d'arbre de décision. Pour les représenter, plusieurs formats sont disponibles, comme le montre l'état de l'art. Toutefois, les formats existants présentent de nombreuses limites, en particulier celle de ne pas exploiter des standards reconnus, ce qui limite leur portabilité, tant au niveau de l'édition que de l'exploitation. Ainsi, le *framework* KCATOS propose un langage de représentation d'arbres de décision simple à comprendre, pour lequel une traduction reposant sur les

technologies du Web sémantique est développée. KcatoS propose en outre un éditeur d'arbres, permettant l'édition collaborative en ligne.

Le troisième apport consiste à concilier dans un même système les approches pour la création des GBPI. En effet, deux approches existent actuellement. La première, s'appuyant sur les connaissances, propose des formats pour les connaissances décisionnelles. La seconde, dite documentaire, considère le GBPI comme un document pouvant être enrichi lors de sa formalisation. Ainsi, KcatoS est un exemple de format de l'approche s'appuyant sur les connaissances tandis qu'OncoLogiK implémente l'approche documentaire. Leur fonctionnement conjoint permet de proposer une solution bénéficiant des avantages des deux approches.

De nombreuses perspectives sont exposées. La plupart d'entre elles visent à améliorer les services aux utilisateurs et l'expressivité de la base de connaissances, notamment en améliorant la qualité des annotations syntaxiques, en renforçant le lien entre KcatoS et OncoLogiK et en intégrant de nouveaux modèles de représentation. En prenant en compte le travail effectué et les perspectives réalisables à moyen terme, un modèle réaliste visant à faire du projet Kasimir un système d'aide à la décision complet est proposé.

Liste des abréviations utilisées

API *Application Programming Interface*

CCAM Classification Commune des Actes Médicaux

CFC Continuum de formalisation des connaissances

CIM Classification statistique internationale des maladies et des problèmes de santé connexes

CP *Content Ontology Design Patterns*

DTD *Document Type Definitions*

FMA *Foundational Model of Anatomy*

G-DEE *Guidelines Document Engineering Environment*

GALEN *Generalized Architecture for Languages Encyclopaedias and Nomenclature*

GBPC Guides de bonnes pratiques cliniques

GBPI Guides de bonnes pratiques informatisés

GEM *Guideline Elements Model*

GLARE *GuideLine Acquisition, Representation, and Execution*

GLIF *Guideline Interchange Format*

GO *Gene Ontology*

HAS Haute Autorité de Santé

IRI *Internationalized Resource Identifiers*

KiWi *Knowledge in Wiki*

LD Logiques de descriptions

MeSH *Medical Subject Headings*

MLM *Medical Logic Modules*

OBO *Open Biomedical Ontologies*

ODP *Ontology Design Patterns*

OfW *Ontologies for Wikis*

OWL *Web Ontology Language*

PRODIGY *Prescribe RatiOnally with Decision-support In General Practice studY*

PTS Portail Terminologique de Santé

RCP Réunion de concertation pluri-disciplinaire

RDF *Resource Description Framework*

RDFa *Resource Description Framework attributes*

RPC Recommandations de pratique clinique

SAD Systèmes d'aide à la décision

SMW *Semantic MediaWiki*

SNOMED *Systematized Nomenclature of Medicine*

SNOMED CT *Systematized Nomenclature of Medicine - Clinical Terms*

SPARQL *SPARQL Protocol and RDF Query Language*

UMLS *Unified Medical Language System*

URI *Uniform Resource Identifiers*

W3C *World Wide Web consortium*

WfO *Wikis for Ontologies*

WYSIWYG *What You See Is What You Get*

Bibliographie

A2ZI (2013). Site Web d'A2ZI. Dernière consultation : Mars 2013. `http://www.a2zi.fr`.

Abidi, S. R. (2007). Ontology-based modeling of breast cancer follow-up clinical practice guideline for providing clinical decision support. *In Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems*, CBMS '07, pages 542–547, Washington, DC, USA. IEEE Computer Society.

Abidi, S. S. et Shayegani, S. (2009). Modeling the form and function of clinical practice guidelines : An ontological model to computerize clinical practice guidelines. *In* Riaño, D., éditeur : *Knowledge Management for Health Care Procedures*, pages 81–91. Springer-Verlag.

AceWiki (2013). Site Web d'AceWiki. Dernière consultation : Mars 2013. `http://attempto.ifi.uzh.ch/acewiki/`.

AHRQ (2013). National Guideline Clearinghouse. Dernière consultation : Mars 2013. `http://www.guideline.gov/`.

Anselma, L., Terenziani, P., Montani, S. et Bottrighi, A. (2007). Applying ai temporal reasoning techniques to clinical guidelines. *In Proceedings of 20th International Joint Conference on Artificial Intelligence (IJCAI-07) Workshop on Spatial and Temporal Reasoning, Hyderabad, India*, pages 1–10.

ANSI (2013). Site Web d'ANSI. Dernière consultation : Mars 2013. `http://www.ansi.org/`.

Appelquist, D., Brickley, D., Carvahlo, M., Iannella, R., Passant, A., Perey, C. et Story, H. (2010). A standards-based, open and privacy-aware social web. *W3C Incubator Group Report, W3C (December 2010)*.

Argüello, M., Des, J., Fernandez-Prieto, M., Perez, R. et Paniagua, H. (2009). Executing medical guidelines on the web : Towards next generation healthcare. *In* Allen, T., Ellis, R. et Petridis, M., éditeurs : *Applications and Innovations in Intelligent Systems XVI*, pages 197–210. Springer.

Auer, S., Dietzold, S., , Riechert, T. et Riechert, T. (2006). Ontowiki - a tool for social, semantic collaboration. *In The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006*, pages 736–749. Springer.

Aumüller, D. et Auer, S. (2005). Towards a Semantic Wiki Experience - Desktop Integration and Interactivity in WikSAR. *In 1st Workshop on The Semantic Desktop, Next Generation Personal Information Management and Collaboration Infrastructure, Galway, Ireland*, pages 388–396.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D. et Patel-Schneider, P. (2003). *The Description Logic Handbook : Theory, Implementation and Applications*. Cambridge University Press.

Bacher, J., Hoehndorf, R. et Kelso, J. (2008). BOWiki : Ontology-based Semantic Wiki with ABox Reasoning. *In 3rd Semantic Wiki Workshop*, volume 360 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Badra, F., Bendaoud, R., Bentebibel, R., Champin, P.-A., Cojan, J., Cordier, A., Desprès, S., Jean-Daubias, S., Lieber, J., Meilender, T., Mille, A., Nauer, E., Napoli, A. et Toussaint, Y. (2008a). TAAABLE : Text Mining, Ontology Engineering, and Hierarchical Classification for Textual Case-Based Cooking. *In* Schaaf, M., éditeur : *ECCBR 2008, The 9th European Conference on Case-Based Reasoning, Trier, Germany, September 1-4, 2008, Workshop Proceedings*, pages 219–228.

Badra, F., d'Aquin, M., Lieber, J. et Meilender, T. (2008b). Edhibou : a customizable interface for decision support in a semantic portal. *In* Bizer, C. et Joshi, A., éditeurs : *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC2008), Karlsruhe, Germany, October 28, 2008*, volume 401 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Baneyx, A. (2007). Construire une ontologie de la Pneumologie. Thèse de Doctorat d'Université, Université Paris 6.

Barsky, E. et Giustini, D. (2007). Introducing Web 2.0 : wikis for health librarians. *Journal of the Canadian Health Libraries Association JCHLA*, 28(4):147–150.

Batsakis, S. et Petrakis, E. G. M. (2011). SOWL : a framework for handling spatio-temporal information in OWL 2.0. *In Proceedings of the 5th international conference on Rule-based reasoning, programming, and applications*, RuleML'2011, pages 242–249. Springer-Verlag.

Baumeister, J., Reutelshoefer, J. et Puppe, F. (2011a). Engineering intelligent systems on the knowledge formalization continuum. *International Journal of Applied Mathematics and Computer Science*, 21(1):27–39.

Baumeister, J., Reutelshoefer, J. et Puppe, F. (2011b). KnowWE : a Semantic Wiki for knowledge engineering. *Applied Intelligence*, 35(3):323–344.

Beale, T. et Heard, S. (2008). openEHR – architecture overview. Rapport technique, The openEHR Foundation. `https://raw.github.com/openEHR/specifications/Release-1.0.1/publishing/architecture/overview.pdf`.

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P. et Morissette, J. (2008). Bio2rdf : Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, 41(5):706–716.

Ben Abbès, S., Scheuermann, A., Meilender, T. et d'Aquin, M. (2012). Characterizing modular ontologies. *In* Schneider, T. et Walther, D., éditeurs : *WoMO*, volume 875 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Bendaoud, R., Napoli, A. et Toussaint, Y. (2008). Formal Concept Analysis : A Unified Framework for Building and Refining Ontologies. *In* Gangemi, A. et Euzenat, J., éditeurs :

Knowledge Engineering : Practice and Patterns, 16th International Conference, EKAW 2008, Acitrezza, Italy, September 29 - October 2, 2008. Proceedings, volume 5268 de *Lecture Notes in Computer Science*, pages 156–171. Springer.

Berners-Lee, T. (2006). Linked data. *W3C Design Issues*. `http://www.w3.org/DesignIssues/LinkedData.html`.

Berners-Lee, T. et Connolly, D. (2008). Notation3 (N3) : A readable RDF syntax. Rapport technique, W3C. `http://www.w3.org/TeamSubmission/n3/`.

Berners-Lee, T. et Fischetti, M. (1999). *Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1st édition.

Berners-Lee, T., Hendler, J. et Lassila, O. (2001). The semantic web. *Scientific American*, 284(5):34–43.

BioPortal (2013). Site Web du projet BioPortal. Dernière consultation : Mars 2013. `http://bioportal.bioontology.org/`.

Birbeck, M. et Adida, B. (2008). RDFa Primer. W3C note, W3C. `http://www.w3.org/TR/2008/NOTE-xhtml-rdfa-primer-20081014/`.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. et Hellmann, S. (2009). Dbpedia - a crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165.

Bodenreider, O. (2008). Biomedical ontologies in action : Role in knowledge management, data integration and decision support. *In AIMIA Yearbook Medical Informatics*, pages 67–79.

Bodenreider, O. et Burgun, A. (2005). Biomedical ontologies. *Medical Informatics : Advances in Knowledge Management and Data Mining in Biomedicine*, pages 211–236.

Bodenreider, O. et Stevens, R. (2006). Bio-ontologies : current trends and future directions. *Briefings in Bioinformatics*, 7(3):256–274.

Boley, H. et Kifer, M. (2009). RIF overview. W3C working draft, W3C. `http://www.w3.org/TR/2009/WD-rif-overview-20091001/`.

Buffa, M., Gandon, F., Erétéo, G., Sander, P. et Faron, C. (2008). SweetWiki : A semantic wiki. *Journal of Web Semantics*, 6(1):84–97.

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A. et Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines ? A framework for improvement. *JAMA*, 282(15):1458–1465.

Campanini, S. E., Castagna, P. et Tazzoli, R. (2004). Platypus Wiki : a Semantic Wiki Wiki Web. *In Semantic Web Applications and Perspectives, Proceedings of 1st Italian Semantic Web Workshop*.

Casteleiro, M. A. et Diz, J. J. D. (2008). Clinical practice guidelines : A case study of combining OWL-S, OWL, and SWRL. *Knowledge-Based Systems*, 21(3):247–255.

Ceccarelli, M., Donatiello, A. et Vitale, D. (2008). Kon3̂ : A clinical decision support system, in oncology environment, based on knowledge management. *In ICTAI (2)*, pages 206–210. IEEE Computer Society.

Ceusters, W., Smith, B. et Goldberg, L. (2005). A terminological and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine*, 44(4):498–507.

Choi, J., Bakken, S., Lussier, Y. A. et Mendonca, E. A. (2006). Improving the human readability of Arden Syntax medical logic modules using a concept-oriented terminology and object-oriented programming expressions. *CIN : Computers, Informatics, Nursing*, 24:220–225.

Choi, J., Currie, L. M., Wang, D. et Bakken, S. (2007). Encoding a clinical practice guideline using guideline interchange format : A case study of a depression screening and management guideline. *International Journal of Medical Informatics*, 76, Supplement 2(0):S302 – S307. Nursing Informatics 2006 Special Issue.

Cimino, J. J. et Zhu, X. (2006). The practical impact of ontologies on biomedical informatics. *Yearbook of Medical Informatics*, pages 124–135.

CISMeF (2013). Catalogue et Index des Sites Médicaux de langue Française. Dernière consultation : Mars 2013. `http://www.chu-rouen.fr/cismef/`.

CiSMeF (2013). Portail Terminologique de Santé. Dernière consultation : Mars 2013. `http://pts.chu-rouen.fr/`.

Clayton, P., Pryor, T., Wigertz, O. et Hripcsak, G. (1989). Issues and Structures for Sharing Medical Knowledge Among Decision-Making Systems : The 1989 Arden Homestead Retreat. *In* Kingsland, L., éditeur : *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 116–21.

Cordier, A., Lieber, J., Molli, P., Nauer, E., Skaf-Molli, H. et Toussaint, Y. (2009). WikiTaaable : A semantic wiki as a blackboard for a textual case-based reasoning system. *In 4th Workshop on Semantic Wikis (SemWiki2009), held in the 6th European Semantic Web Conference*, pages 18–32.

CosmoCode (2013). Site Web WikiMatrix. Dernière consultation : Mars 2013. `http://www.wikimatrix.org/`.

Cossac (2013). Page Web de présentation de Tallis. Dernière consultation : Mars 2013. `http://www.cossac.org/tallis`.

Cunningham, W. (2002). What Is Wiki. Dernière consultation : Mars 2013. `http://www.wiki.org/wiki.cgi?WhatIsWiki`.

d3web (2013). Site Web de d3web. Dernière consultation : Mars 2013. `http://www.d3web.de/`.

d'Aquin, M. (2005). Un portail sémantique pour la gestion des connaissances en cancérologie. Thèse de Doctorat d'Université, Université Henri Poincaré - Nancy 1.

d'Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A. et Szathmary, L. (2007). Case base mining for adaptation knowledge acquisition. *In* Veloso, M. M., éditeur : *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 750–755.

d'Aquin, M., Bouthier, C., Brachais, S., Lieber, J. et Napoli, A. (2005). Knowledge editing and maintenance tools for a semantic portal in oncology. *International Journal on Human-Computer Studies*, 62(5):619–638.

d'Aquin, M., Lieber, J. et Napoli, A. (2006a). Adaptation knowledge acquisition : A case study for case-based decision support in oncology. *Computational Intelligence*, 22(3-4):161–176.

d'Aquin, M., Lieber, J. et Napoli, A. (2006b). Towards a Semantic Portal for Oncology using a Description Logic with Fuzzy Concrete Domains. *In* Elie Sanchez, éditeur : *Fuzzy Logic and the Semantic Web*, volume 1 de *Capturing intelligence*. Elsevier.

Darmoni, S. J., Thirion, B., Leroy, J. P., Douyère, M., Baudic, F. et Piot, J. (2000). CISMeF : a structured health resource guide for healthcare professionals and patients. *In* Mariani, J.-J. et Harman, D., éditeurs : *Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications) - RIAO 2000, 6th International Conference, College de France, France, April 12-14, 2000. Proceedings*, pages 819–829. CID.

Dassen, W., Gorgels, A., Berendsen, A., Dijk, W., de Clercq, P., Hasman, A. et Baljon, M. (2003). Guideline assessment and implementation in congestive heart failure. *In Computers in Cardiology*, pages 331 – 334.

DBpedia (2013). Site Web du projet DBpedia. Dernière consultation : Mars 2013. `http://dbpedia.org/`.

DBpedia.fr (2012). Site Web de la version francophone du projet DBpedia. `http://fr.dbpedia.org/`.

DCMI (2013). Dcmi metadata terms. Dernière consultation : Mars 2013. `http://www.dublincore.org/documents/dcmi-terms`.

de Clercq, P., Kaiser, K. et Hasman, A. (2008). Computer-interpretable guideline formalisms. *In* ten Teije, A., Miksch, S. et Lucas, P., éditeurs : *Computer-based Medical Guidelines and Protocols : A Primer and Current Trends*, pages 22–43. IOS Press.

de Clercq, P. A., Blom, J. A., Korsten, H. H. et Hasman, A. (2004). Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine*, 31(1):1 – 27.

de Clercq, P. A., Hasman, A., Blom, J. A. et Korsten, H. H. (2001). Design and implementation of a framework to support the development of clinical guidelines. *International Journal of Medical Informatics*, 64:285 – 318.

de Coi, J. L., Fuchs, N. E., Kaljurand, K. et Kuhn, T. (2009). Controlled English for reasoning on the Semantic Web. *In* Bry, F. et Małuszyński, J., éditeurs : *Semantic Techniques for the Web — The REWERSE Perspective*, volume 5500 de *Lecture Notes in Computer Science*, pages 276–308. Springer.

Dean, M., Connolly, D., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F. et Stein, L. A. (2002). OWL Web Ontology Language 1.0 Reference. Rapport technique, W3C Working Draft.

DIQA (2013). Présentation de DataWiki. Dernière consultation : Mars 2013. `http://diqa-pm.com/en/DataWiki`.

Doucette, J. A., Khan, A. et Cohen, R. (2012). A comparative evaluation of an ontological medical decision support system (omed) for critical environments. *In* Luo, G., Liu, J. et Yang, C. C., éditeurs : *IHI*, pages 703–708. ACM.

Dufour-Lussier, V., Ber, F. L., Lieber, J., Meilender, T. et Nauer, E. (2012). Semi-automatic annotation process for procedural texts : An application on cooking recipes. *In Cooking with Computers (CwC), Workshop at ECAI 2012, Montpellier, France, August 28, 2012*.

Epperson, B., Ishikawa, M., Dubinko, M., Pemberton, S., McCarron, S., Axelsson, J., Navarro, A. et Birbeck, M. (2006). XHTML^{TM} 2.0. W3C working draft, W3C. `http://www.w3.org/TR/2006/WD-xhtml2-20060726`.

Erinoff, E. G. (2011). Feasibility study of a wiki collaboration platform for systematic reviews. *Methods Research Report*. AHRQ Publication No. 11 EHC040 EF. Rockville, MD : Agency for Healthcare Research and Quality. Available at : `www.effectivehealthcare.ahrq.gov/reports/final.cfm`.

Feigenbaum, E. A. et McCorduck, P. (1983). *The fifth generation - artificial intelligence and Japan's computer challenge to the world*. Addison-Wesley.

Feigenbaum, L. et Johnston, S. (2009). SPARQL 1.1 protocol for RDF. W3C working draft, W3C. `http://www.w3.org/TR/2009/WD-sparql11-protocol-20091022/`.

Fischer, J., Gantner, Z., Rendle, S., Stritt, M. et Schmidt-Thieme, L. (2006). Ideas and improvements for semantic wikis. *In* Sure, Y. et Domingue, J., éditeurs : *3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006*, pages 650–663. Springer.

Fox, J., Johns, N., Lyons, C., Rahmanzadeh, A., Thomson, R. et Wilson, P. (1997). Proforma : a general technology for clinical decision support systems. *Computer Methods and Programs in Biomedicine*, 54(1-2):59 – 67.

Freitas, F., Schulz, S. et Moraes, E. (2009). Survey of current terminologies and ontologies in biology and medicine. *In RECIIS. Electronic Journal of Communication, Information and Innovation in Health (English edition. Online), v. 3*, pages 1–13.

Fung, K. W. et Bodenreider, O. (2012). Knowledge Representation and Ontologies. *Health Informatics, Clinical Research Informatics*, 3:255–275.

Gangemi, A. (2005). Ontology Design Patterns for Semantic Web Content. *In* Gil, Y., Motta, E., Benjamins, V. R. et Musen, M. A., éditeurs : *The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, volume 3729 de *Lecture Notes in Computer Science*, pages 262–276. Springer.

Gangemi, A. et Presutti, V. (2009). Ontology Design Patterns. *In* Staab, S. et Studer, R., éditeurs : *Handbook on Ontologies*, International Handbooks on Information Systems, pages 221–243. Springer.

Gene Ontology (2013). Site Web de Gene Ontology. Dernière consultation : Mars 2013. `http://www.geneontology.org/`.

Georg, G. (2006). Analyse informatique de guides de bonnes pratiques cliniques. Thèse de Doctorat d'Université, Université Paris 6.

Georg, G. et Jaulent, M. C. (2005). An environment for document engineering of clinical guidelines. *AMIA Annual Symposium Proceedings*, pages 276–280.

Glasziou, P., Burls, A. et Gilbert, R. (2008). Evidence based medicine and the medical curriculum. *BMJ*, 337(sep24 3).

Glimm, B., Horrocks, I., Motik, B. et Stoilos, G. (2010). Optimising ontology classification. *In* Patel-Schneider, P., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J., Horrocks, I. et Glimm, B., éditeurs : *The Semantic Web - ISWC 2010*, volume 6496 de *Lecture Notes in Computer Science*, pages 225–240. Springer.

Google (2013a). Google Scholar. Dernière consultation : Mars 2013. `http://scholar.google.fr/`.

Google (2013b). Google Web Toolkit. Dernière consultation : Mars 2013. `http://code.google.com/intl/en/webtoolkit/`.

Gordon, C., Veloso, M. et The PRESTIGE Consortium (1999). Guidelines in healthcare : the experience of the Prestige project. *Studies in health technology and informatics*, 68:733–738.

Grando, M. A., Glasspool, D. et Fox, J. (2012). A formal approach to the analysis of clinical computer-interpretable guideline modeling languages. *Artificial Intelligence in Medicine*, 54(1): 1–13.

Grosjean, J., Merabti, T., Dahamna, B., Kergourlay, I., Thirion, B., Soualmia, L. F. et Darmoni, S. J. (2011). Health multi-terminology portal : a semantic added-value for patient safety. *Studies in Health Technology and Informatics*, 166:129–138.

Gruber, T. (2008). Collective knowledge systems : Where the Social Web meets the Semantic Web. *Journal of Web Semantics*, 6(1):4–13.

Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *In International Journal of Human-Computer Studies*, pages 907–928. Kluwer Academic Publishers.

Haake, A., Lukosch, S. et Schümmer, T. (2005). Wiki-templates : adding structure support to wikis on demand. *In Proceedings of the 2005 international symposium on Wikis*, WikiSym '05, pages 41–51, New York, NY, USA. ACM.

Hagerty, C. G., Pickens, D., Kulikowski, C. et Sonnenberg, F. (2000). HGML : a hypertext guideline markup language. *Proceedings of AMIA Symposium*, pages 325–329.

Hatko, R., Reutelshoefer, J., Baumeister, J. et Puppe, F. (2010). Modeling of Diagnostic Guideline Knowledge in Semantic Wikis. *In Proceedings of the Workshop on Open Knowledge Models (OKM-2010) at the 17th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*.

Hatko, R., Schädler, D., Mersmann, S., Baumeister, J., Weiler, N. et Puppe, F. (2012). Implementing an Automated Ventilation Guideline using the Semantic Wiki KnowWE. *In* Stuckenschmidt, H., ten Teije, A. et Voelker, J., éditeurs : *EKAW 2012 : 18th International Conference on Knowledge Engineering and Knowledge Management*.

Haute Autorité de la Santé (2010). Recommandations pour la pratique clinique. Dernière consultation : Mars 2013. http://www.has-sante.fr/portail/jcms/c_431294.

He, S., Nachimuthu, S. K., Shakib, S. C. et Lau, L. M. (2009). Collaborative authoring of biomedical terminologies using a semantic wiki. *AMIA Annual Symposium Proceedings*, 2009:234–8.

Heino, N., Dietzold, S., Martin, M. et Auer, S. (2009). Developing Semantic Web Applications with the OntoWiki Framework. *In* Pellegrini, T., Auer, S., Tochtermann, K. et Schaffert, S., éditeurs : *Networked Knowledge - Networked Media*, volume 221 de *Studies in Computational Intelligence*, pages 61–77. Springer.

Herzig, D. M. et Ell, B. (2010). Semantic Mediawiki in operation : experiences with building a semantic portal. *In Proceedings of the 9th international semantic web conference on The semantic web - Volume Part II*, ISWC'10, pages 114–128. Springer-Verlag.

HL7 (2013a). Page Web de présentation du standard RIM. Dernière consultation : Mars 2013. http://www.hl7.org/implement/standards/rim.cfm.

HL7 (2013b). Site Web de HL7. Dernière consultation : Mars 2013. http://www.hl7.org/.

Hollingsworth, D. (1995). The Workflow Reference Model. Rapport technique, Workflow Management Coalition.

HonCODE (2013). Site Web du projet HonCODE. Dernière consultation : Mars 2013. http://www.hon.ch.

Horridge, M. et Bechhofer, S. (2011). The OWL API : A Java API for OWL ontologies. *Semantic Web journal*, 2(1):11–21.

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B. et Dean, M. (2004). SWRL : A Semantic Web Rule Language Combining OWL and RuleML. W3c member submission, World Wide Web Consortium. http://www.w3.org/Submission/SWRL.

Hoyt, D. B. (1997). Clinical practice guidelines. *The American Journal of Surgery*, 173(1):32–34.

Huang, M., Névéol, A. et Lu, Z. (2011). Recommending mesh terms for annotating biomedical articles. *Journal of the American Medical Informatics Association : JAMIA*, 18(5):660–667.

Hussain, S., Abidi, S. R. et Abidi, S. S. R. (2007). Semantic web framework for knowledge-centric clinical decision support systems. *In* Bellazzi, R., Abu-Hanna, A. et Hunter, J., éditeurs : *AIME*, volume 4594 de *Lecture Notes in Computer Science*, pages 451–455. Springer.

IHTSDO (2013). Page Web de présentation de SNOMED CT. Dernière consultation : Mars 2013. http://www.ihtsdo.org/snomed-ct/.

Infermed (2013). Page Web de présentation d'Arezzo. Dernière consultation : Mars 2013. http://www.infermed.com/index.php/arezzo/Arezzo_Technology.

Intute (2013). Intute. Dernière consultation : Mars 2013. http://www.intute.ac.uk.

Isern, D. et Moreno, A. (2008). Computer-based execution of clinical guidelines : A review. *International Journal of Medical Informatics*, 77(12):787 – 808.

Jacobs, A., Quinn, T. et Nelson, S. (2006). Mapping SNOMED-CT concepts to MeSH concepts. *AMIA Annual Symposium Proceedings*.

Jafarpour, B., Abidi, S. R. et Abidi, S. S. R. (2011). Exploiting OWL reasoning services to execute ontologically-modeled clinical practice guidelines. *In Proceedings of the 13th conference on Artificial intelligence in medicine*, AIME'11, pages 307–311. Springer-Verlag.

Jenders, R. A., Corman, R. et Dasgupta, B. (2003). Making the standard more standard : a data and query model for knowledge representation in the Arden syntax. *AMIA Annual Symposium Proceedings*, pages 323–330.

Jiang, G., Solbrig, H. R., Iberson-Hurst, D., Kush, R. D. et Chute, C. G. (2010). A Collaborative Framework for Representation and Harmonization of Clinical Study Data Elements Using Semantic MediaWiki. *AMIA Summits Transl Sci Proc*, 2010:11–15.

Johnson, P., Tu, S. et Jones, N. (2001). Achieving reuse of computable guideline systems. *Studies in health technology and informatics*, 84(Pt 1):99–103.

Kaiser, K. et Miksch, S. (2005). Modeling computer-supported clinical guidelines and protocols. Rapport technique, Vienna University of Technology.

KAON2 (2013). Page Web de présentation de KAON2. Dernière consultation : Mars 2013. `http://kaon2.semanticweb.org`.

Kashyap, V., Morales, A. et Hongsermeier, T. (2005). Creation and Maintenance of Implementable Clinical Guideline Specifications. Rapport technique, Partners HealthCare.

Kasimir (2013). Site Web du projet Kasimir. Dernière consultation : Mars 2013. `http://kasimir.loria.fr`.

Kiesel, M. (2006). Kaukolu : Hub of the semantic corporate intranet. *In* Völkel, M. et Schaffert, S., éditeurs : *SemWiki2006, First Workshop on Semantic Wikis - From Wiki to Semantics, Proceedings, co-located with the ESWC2006, Budva, Montenegro, June 12, 2006*. CEUR-WS.org.

Kim, S., Haug, P. J., Rocha, R. A. et Choi, I. (2008). Modeling the Arden Syntax for medical decisions in XML. *International Journal of Medical Informatics*, 77(10):650 – 656.

KiWi (2012). Site Web du projet KiWi. Dernière consultation : Mars 2013. `http://www.kiwi-project.eu/`.

Knoodl (2013). Site Web de Knoodl. Dernière consultation : Mars 2013. `http://knoodl.com`.

Kosara, R. et Miksch, S. (2001). Metaphors of movement : a visualization and user interface for time-oriented, skeletal plans. *Artificial Intelligence in Medicine*, 22(2):111–131.

Krötzsch, M., Patel-Schneider, P. F., Rudolph, S., Hitzler, P. et Parsia, B. (2009). OWL 2 Web Ontology Language Primer. Rapport technique, W3C. `http://www.w3.org/TR/2009/REC-owl2-primer-20091027/`.

Krötzsch, M., Schaffert, S. et Vrandecic, D. (2007a). Reasoning in semantic wikis. *In* Antoniou, G., Assmann, U., Baroglio, C., Decker, S., Henze, N., Patranjan, P.-L. et Tolksdorf, R., éditeurs : *Reasoning Web, Third International Summer School 2007, Dresden,*

Germany, September 3-7, 2007, Tutorial Lectures, volume 4636 de *Lecture Notes in Computer Science*, pages 310–329. Springer.

Krötzsch, M. et Vrandecic, D. (2011). Semantic Mediawiki. *In Foundations for the Web of Information and Services*, pages 311–326.

Krötzsch, M., Vrandecic, D. et Völkel, M. (2005). Wikipedia and the Semantic Web - The Missing Links. *In Proceedings of the WikiMania2005, 4-8 Aug., 2005.*

Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H. et Studer, R. (2007b). Semantic Wikipedia. *Journal of Web Semantics*, 5(4):251–261.

Kuhn, T. (2009). How controlled english can improve semantic wikis. *In* Lange, C., Schaffert, S., Skaf-Molli, H. et Völkel, M., éditeurs : *4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece, June 1st, 2009. Proceedings*, volume 464 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Lamy, J.-B., Ebrahiminia, V., Riou, C., Seroussi, B., Bouaud, J., Simon, C., Dubois, S., Butti, A., Simon, G., Favre, M., Falcoff, H. et Venot, A. (2010). How to translate therapeutic recommendations in clinical practice guidelines into rules for critiquing physician prescriptions ? methods and application to five guidelines. *BMC Medical Informatics and Decision Making*, 10(1):31.

Lange, C. (2008). Mathematical semantic markup in a wiki : The roles of symbols and notations. *In 3rd Semantic Wiki Workshop*, volume 360 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Latoszek-Berendsen, A., Tange, H., van den Herik, H. J. et Hasman, A. (2010). From clinical practice guidelines to computer-interpretable guidelines. A literature overview. *Methods of information in medicine*, 49(6):550–570.

Leong, T. Y., Kaiser, K. et Miksch, S. (2007). Free and open source enabling technologies for patient-centric, guideline-based clinical decision support : a survey. *Yearbook of medical informatics*, pages 74–86.

Leuf, B. et Cunningham, W. (2001). *The Wiki way : quick collaboration on the Web*. Addison-Wesley.

Lieber, J. (2006). Contributions à la conception de systèmes de raisonnement à partir de cas. Habilitation à Diriger des Recherches, Université Henri Poincaré - Nancy 1.

Lieber, J., d'Aquin, M., Badra, F. et Napoli, A. (2008). Modeling adaptation of breast cancer treatment decision protocols in the kasimir project. *Applied Intelligence*, 28(3):261–274.

Lokker, C., Haynes, R. B., Wilczynski, N. L., McKibbon, K. A. et Walter, S. D. (2011). Retrieval of diagnostic and treatment studies for clinical use through PubMed and PubMed's Clinical Queries filters. *Journal of the American Medical Informatics Association*, 18(5):652–659.

Louridas, P. (2006). Using wikis in software development. *IEEE Software*, 23:88–91.

Lugtenberg, M., Burgers, J., Besters, C., Han, D. et Westert, G. (2011). Perceived barriers to guideline adherence : A survey among general practitioners. *BMC Family Practice*, 12(1):98.

Mader, S. (2007). *Wikipatterns*. Wiley, 1 édition.

Maler, E., Paoli, J., Sperberg-McQueen, C. M., Yergeau, F. et Bray, T. (2004). Extensible Markup Language (XML) 1.0 (Third Edition). first edition of a Recommendation, W3C. `http://www.w3.org/TR/2004/REC-xml-20040204`.

Martin, M., Gerber, D., Heino, N., Auer, S. et Ermilov, T. (2011). Managing multimodal and multilingual semantic content. *In Proceedings of the 7th International Conference on Web Information Systems and Technologies*.

MediaWiki (2013). Site Web de MediaWiki. Dernière consultation : Mars 2013. `http://www.mediawiki.org/`.

Meilender, T. (2012a). KCATOS : Documentation technique. Rapport interne, A2ZI.

Meilender, T. (2012b). OncoLogiK : Documentation technique. Rapport interne, A2ZI.

Meilender, T., Jay, N., Lieber, J. et Palomares, F. (2011a). Édition sémantique d'arbres de décision pour l'oncologie avec KcatoS. *In* Tamine, L., Darmoni, S. et Soualmia, L., éditeurs : *Actes de SIIM 2011, 1er Symposium sur l'Ingénierie de l'Information Médicale*, http ://www.irit.fr. IRIT Press.

Meilender, T., Jay, N., Lieber, J. et Palomares, F. (2011b). Les moteurs de wikis sémantiques : un état de l'art. *In* Khenchaf, A. et Poncelet, P., éditeurs : *Extraction et gestion des connaissances (EGC'2011), Actes, 25 au 29 janvier 2011, Brest, France*, volume RNTI-E-20 de *Revue des Nouvelles Technologies de l'Information*, pages 575–580. Hermann-Éditions.

Meilender, T., Lieber, J., Palomares, F., Herengt, G. et Jay, N. (2012a). A semantic wiki for editing and sharing decision guidelines in oncology. *In* Mantas, J., Andersen, S. K., Mazzoleni, M. C., Blobel, B., Quaglini, S. et Moen, A., éditeurs : *Quality of Life through Quality of Information*, volume 180 de *Studies in health technology and informatics*, pages 411–5. IOS Press.

Meilender, T., Lieber, J., Palomares, F. et Jay, N. (2012b). From web 1.0 to social semantic web : Lessons learnt from a migration to a medical semantic wiki. *In* Simperl, E., Cimiano, P., Polleres, A., Corcho, Ó. et Presutti, V., éditeurs : *The Semantic Web : Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceeding*, volume 7295 de *Lecture Notes in Computer Science*, pages 618–632. Springer.

Meilender, T., Lieber, J., Palomares, F. et Jay, N. (2012c). Semantic decision trees editing for decision support with KCATOS - Application in oncology. *In ESWC Workshops, SIMI 2012 : Semantic Interoperability in Medical Informatics, Heraklion, Greece, May 27-28, 2012*.

Miller, E. et Manola, F. (2004). RDF primer. W3C Recommendation, W3C. `http://www.w3.org/TR/2004/REC-rdf-primer-20040210/`.

Mishra, R. B. et Kumar, S. (2011). Semantic web reasoners and languages. *Artificial Intelligence Review*, 35(4):339–368.

Montori, V. M., Wilczynski, N. L., Morgan, D. et Haynes, R. B. (2005). Optimal search strategies for retrieving systematic reviews from Medline : analytical survey. *BMJ*, 330(7482): 68.

Motik, B., Parsia, B. et Patel-Schneider, P. F. (2009). OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. W3C Recommendation, W3C. http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/.

Mulyar, N., Van Der Aalst, W. M. P. et Peleg, M. (2007). A pattern-based analysis of clinical computer-interpretable guideline modeling languages. *Journal of the American Medical Informatics Association : JAMIA*, 14(6):781–787.

Musen, M. A., Tu, S. W., Das, A. K. et Shahar, Y. (1996). Eon : A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association : JAMIA*, 3(6):367–388.

NCBI (2013a). Éditeur de requêtes MeSH. Dernière consultation : Mars 2013. http://www.ncbi.nlm.nih.gov/mesh/68013274?report=Full.

NCBI (2013b). PubMed. Dernière consultation : Mars 2013. http://www.ncbi.nlm.nih.gov/pubmed/.

NCBI (2013c). Site Web de présentation du MeSH. Dernière consultation : Mars 2013. http://www.ncbi.nlm.nih.gov/mesh/.

NCI (2013). Page Web de présentation du NCI Thesaurus. Dernière consultation : Mars 2013. http://ncit.nci.nih.gov/.

NICE (2013). NHS Evidence. Dernière consultation : Mars 2013. http://www.evidence.nhs.uk.

NLM (2013a). Page web de présentation de MEDLINE. Dernière consultation : Mars 2013. http://www.nlm.nih.gov/bsd/pmresources.html.

NLM (2013b). Page Web de présentation d'UMLS. Dernière consultation : Mars 2013. http://www.nlm.nih.gov/research/umls/.

NN/LM (2013a). Qualifiers - 2013. Dernière consultation : Mars 2013. http://www.nlm.nih.gov/mesh/topsubscope.html.

NN/LM (2013b). Searching PubMed with MeSH. Dernière consultation : Mars 2013. http://nnlm.gov/training/ressources/meshtri.pdf.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A. D., Chute, C. G. et Musen, M. A. (2009). Bioportal : ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37(Web-Server-Issue):170–173.

OBO Foundry (2013). Site Web OBO Foundry. Dernière consultation : Mars 2013. http://www.obofoundry.org.

Oncolor (2013). Site Web du réseau de santé Oncolor. Dernière consultation : Mars 2013. http://www.oncolor.org.

Ongenae, F., Backere, F., Steurbaut, K., Colpaert, K., Kerckhove, W., Decruyenaere, J. et Turck, F. (2010). Towards computerizing intensive care sedation guidelines : design of a rule-based architecture for automated execution of clinical guidelines. *BMC Medical Informatics and Decision Making*, 10:1–22.

OntologyDesignPatterns.org (2010). N-Ary Relation Pattern (OWL 2). Dernière consultation : Mars 2013. http://ontologydesignpatterns.org/wiki/Submissions:N-Ary_Relation_Pattern_(OWL_2).

OntoWiki (2012). Site Web d'OntoWiki. Dernière consultation : Mars 2013. http://ontowiki.net.

OpenClinical (2013). Site Web communautaire OpenClinical. Dernière consultation : Mars 2013. http://www.openclinical.org.

OpenGalen (2013). Site Web du projet OpenGalen. Dernière consultation : Mars 2013. http://www.opengalen.org/.

O'Reilly, T. (2005). O'Reilly Network : What Is Web 2.0.

Orpailleur (2013). Site Web de l'équipe Orpailleur. Dernière consultation : Mars 2013. http://orpailleur.loria.fr.

Pan, F. et Hobbs, J. R. (2006). Time Ontology in OWL. W3C working draft, W3C. http://www.w3.org/TR/2006/WD-owl-time-20060927/.

Papakonstantinou, D., Malamateniou, F. et Vassilacopoulos, G. (2011). A semantic wiki for user training in ePrescribing processes. *In Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '11, pages 30 :1–30 :6. ACM.

Parker, K. R. et Chao, J. T. (2007). Wiki as a Teaching Tool. *Interdisciplinary Journal of Knowledge and Learning Objects*, 3.

Peleg, M., Boxwala, A. A., Bernstam, E., Tu, S., Greenes, R. A. et Shortliffe, E. H. (2001). Sharable representation of clinical guidelines in GLIF : relationship to the Arden Syntax. *Journal of Biomedical Informatics*, 34:170–181.

Peleg, M., Boxwala, A. A., Ogunyemi, O. et Greenes, R. A. (2000). GLIF3 : the evolution of a guideline representation format. *In Proceedings of the AMIA Symposium*, pages 645–649.

Peleg, M., Fodor, D., Keren, A. et Karnieli, S. (2006). Adaptation of Practice Guidelines for Clinical Decision Support : A Case Study of Diabetic Foot Care. *In Proceeding of the the biennial European Conference on Artificial Intelligence (ECAI) 2006, Workshop : AI techniques in healthcare : computerized guidelines and protocols, Riva del Garda, Italy.*

Peleg, M., Keren, S. et Denekamp, Y. (2008). Mapping computerized clinical guidelines to electronic medical records : Knowledge-data ontological mapper (KDOM). *Journal of Biomedical Informatics*, 41(1):180–201.

Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R. A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E. H., Stefanelli, M. et et al. (2003). Comparing computer-interpretable guideline models : A case-study approach. *Journal of the American Medical Informatics Association : JAMIA*, 10:2003.

Peleg, M., Tu, S. W., Leonardi, G., Quaglini, S., Russo, P., Palladini, G. et Merlini, G. (2012). Reasoning with effects of clinical guideline actions using owl : Al amyloidosis as a case study. *In Proceedings of the 3rd international conference on Knowledge Representation for Health-Care*, KR4HC'11, pages 65–79. Springer-Verlag.

Pereira, S., Neveol, A., Kerdelhué, G., Serrot, E., Joubert, M. et Darmoni, S. J. (2008). Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. *AMIA Annual Symposium Proceedings*, pages 586–590.

PRODIGY (2011). Hypothyroidism. Dernière consultation : Mars 2013. `http://www.prodigy.clarity.co.uk/hypothyroidism`.

Prud'hommeaux, E. et Seaborne, A. (2008a). SPARQL Query Language for RDF. W3C Recommendation. `http://www.w3.org/TR/rdf-sparql-query/`.

Prud'hommeaux, E. et Seaborne, A. (2008b). SPARQL query language for RDF. W3C Recommendation, W3C. `http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/`.

Quaglini, S., Stefanelli, M., Cavallini, A., Micieli, G., Fassino, C. et Mossa, C. (2000). Guideline-based careflow systems. *Artificial Intelligence in Medicine*, 20(1):5–22.

Rakebul, H. (2009). Clip-moki : A collaborative tool for the modeling of clinical guideline. Thèse de Doctorat d'Université, Università degli Studi di Trento.

Reutelshöfer, J., Haupt, F., Lemmerich, F. et Baumeister, J. (2009). An extensible semantic wiki architecture. *In* Lange, C., Schaffert, S., Skaf-Molli, H. et Völkel, M., éditeurs : *4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece, June 1st, 2009. Proceedings*, volume 464 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Riechert, T., Morgenstern, U., Auer, S., Tramp, S. et Martin, M. (2010). Knowledge engineering for historians on the example of the catalogus professorum lipsiensis. *In* Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J. Z., Horrocks, I. et Glimm, B., éditeurs : *Proceedings of the 9th International Semantic Web Conference (ISWC2010)*, volume 6497 de *Lecture Notes in Computer Science*, pages 225–240, Shanghai / China. Springer.

Russell, N., Hofstede, A. H. M. T. et Mulyar, N. (2006). Workflow ControlFlow Patterns : A Revised View. Rapport technique, BPM Center Report BPM-06-22 , BPMcenter.org.

SAGE (2013). Project overview. Dernière consultation : Mars 2013. `http://sage.wherever.org/`.

Sailors, R. M. et Jenders, R. A. (2002). Arden Syntax 2.1 : Ten Years of Standardized Knowledge Bases. *Proceedings of the AMIA Symposium*.

Schaffert, S. (2006). Ikewiki : A semantic wiki for collaborative knowledge management. *In 15th IEEE International Workshops on Enabling Technologies : Infrastructures for Collaborative Enterprises (WETICE 2006), 26-28 June 2006, Manchester, United Kingdom*, pages 388–396. IEEE Computer Society.

Schatten, M., Čubrilo, M. et Ševa, J. (2009). A Semantic Wiki System Based on F-Logic. *In 19th Central European Conference on Information and Intelligent Systems, Varaždin, Croatia*.

Schreiber, G. et Dean, M. (2004). OWL Web Ontology Language Reference. W3C Recommendation, W3C. `http://www.w3.org/TR/2004/REC-owl-ref-20040210/`.

Semantic MediaWiki (2013). Site Web de Semantic Mediawiki. Dernière consultation : Mars 2013. http://semantic-mediawiki.org/.

Semanticweb.org (2011). Semantic Web standards. Dernière consultation : Mars 2013. http://semanticweb.org/wiki/Semantic_Web_standards.

Semanticweb.org (2012). Site Web Semanticweb.org. Dernière consultation : Mars 2013. http://www.semanticweb.org/.

Semanticweb.org (2013). Présentation de SMW+. Dernière consultation : Mars 2013. http://semanticweb.org/wiki/SMW\%2B.

Shahar, Y., Miksch, S. et Johnson, P. (1996). An intention-based language for representing clinical guidelines. *Proceedings of the AMIA Annual Fall Symposium*, pages 592–596.

Shahar, Y., Miksch, S. et Johnson, P. (1998). The Asgaard project : a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*, 14(1-2):29 – 51.

Shahar, Y., Young, O., Shalom, E., Mayaffit, A., Moskovitch, R., Hessing, A. et Galperin, M. (2003). Degel : A hybrid, multiple-ontology framework for specification and retrieval of clinical guidelines. *In* Dojat, M., Keravnou, E. et Barahona, P., éditeurs : *Artificial Intelligence in Medicine*, volume 2780 de *Lecture Notes in Computer Science*, pages 122–131. Springer.

Shiffman, R. N., Karras, B. T., Agrawal, A., Chen, R., Marenco, L. et Nath, S. (2000). GEM : a proposal for a more comprehensive guideline document model using XML. *Journal of the American Medical Informatics Association : JAMIA*, 7(5):488–498.

Sint, R., Stroka, S., Schaffert, S. et Ferstl, R. (2009). Combining Unstructured, Flly Structured and Semi-Structured Information in Semantic Wikis. *In* Lange, C., Schaffert, S., Skaf-Molli, H. et Völkel, M., éditeurs : *4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece, June 1st, 2009. Proceedings*, volume 464 de *CEUR Workshop Proceedings*. CEUR-WS.org.

Sirin, E., Parsia, B., Grau, B., Kalyanpur, A. et Katz, Y. (2007a). Pellet : A practical OWL-DL reasoner. *Web Semantics : Science, Services and Agents on the World Wide Web*, 5(2):51–53.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A. et Katz, Y. (2007b). Pellet : A practical owl-dl reasoner. *Journal of Web Semantics*, 5(2):51–53.

Skonetzki, S., Gausepohl, H. J., van der Haak, M., Knaebel, S., Linderkamp, O. et Wetter, T. (2004). HELEN, a modular framework for representing and implementing clinical practice guidelines. *Methods of Information in Medicine*, 43(4):413–426.

Smart, P. R., Bao, J., Braines, D. et Shadbolt, N. R. (2009). Development of a Controlled Natural Language Interface for Semantic Mediawiki. *In* Fuchs, N. E., éditeur : *Controlled Natural Language, Workshop on Controlled Natural Language, CNL 2009, Marettimo Island, Italy, June 8-10, 2009. Revised Papers*, volume 5972 de *Lecture Notes in Computer Science*, pages 206–225. Springer.

Sonnenberg, F. et Hagerty, C. (2006). Computer-interpretable clinical practice guidelines. Where are we and where are we going ? *IMIA Yearbook 2006 : Assessing Information - Technologies for Health*, pages 145–158.

Sonntag, D., Setz, J., Baker, M. A. et Zillner, S. (2012). Clinical Trial and Disease Search with Ad Hoc Interactive Ontology Alignments. *In* Simperl, E., Cimiano, P., Polleres, A., Corcho, Ó. et Presutti, V., éditeurs : *The Semantic Web : Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings*, volume 7295 de *Lecture Notes in Computer Science*, pages 674–686. Springer.

Souzis, A. (2006). Bringing the "Wiki-Way" to the Semantic Web with Rhizome. *In* Völkel, M. et Schaffert, S., éditeurs : *SemWiki2006, First Workshop on Semantic Wikis - From Wiki to Semantics, Proceedings, co-located with the ESWC2006, Budva, Montenegro, June 12, 2006*, pages 222–229. CEUR-WS.org.

Stanford (2013). Site Web de Protégé. Dernière consultation : Mars 2013. `http://protege.stanford.edu/`.

Stantic, B., Terenziani, P., Governatori, G., Bottrighi, A. et Sattar, A. (2012). An implicit approach to deal with periodically repeated medical data. *Artificial Intelligence in Medicine*, 55(3):149–162.

Starren, J., Hripcsak, G., Jordan, D., Allen, B., Weissman, C. et Clayton, P. (1994). Encoding a post-operative coronary artery bypass surgery care plan in the Arden Syntax. *Computers in Biology and Medicine*, 24(5):411 – 417.

Subleme (2013). Hébergement du projet Subleme. Dernière consultation : Mars 2013. `http://code.google.com/p/subleme/`.

Swick, R. R. et Lassila, O. (1999). Resource Description Framework (RDF) Model and Syntax Specification. supersed work, W3C. `http://www.w3.org/TR/1999/REC-rdf-syntax-19990222`.

Terenziani, P., Montani, S., Bottrighi, A., Molino, G. et Torchio, M. (2003). Applying artificial intelligence to clinical guidelines : The glare approach. *In AI*IA 2003 : Advances in Artificial Intelligence, 8th Congress of the Italian Association for Artificial Intelligence.* SpringerVerlag.

The ASGAARD Project (2013). Site Web du projet ASGAARD. Dernière consultation : Mars 2013. `http://www.asgaard.tuwien.ac.at/`.

Tramp, S., Frischmuth, P. et Heino, N. (2010). OntoWiki – a Semantic Data Wiki Enabling the Collaborative Creation and (Linked Data) Publication of RDF Knowledge Bases. *In* Corcho, O. et Voelker, J., éditeurs : *Demo Proceedings of the EKAW 2010.*

Tran, N., Michel, G., Krauthammer, M. et Shiffman, R. N. (2009). Embedding the guideline elements model in web ontology language. *AMIA Annual Symposium Proceedings*, 2009:645–649.

Tu, S. W., Campbell, J. et Musen, M. A. (2004). The SAGE guideline modeling : motivation and methodology. *Studies in health technology and informatics*, 101:167–171.

Tu, S. W., Campbell, J. R., Glasgow, J., Nyman, M. A., McClure, R., McClay, J., Parker, C., Hrabak, K. M., Berg, D., Weida, T., Mansfield, J. G., Musen, M. A. et Abarbanel, R. M. (2007). The SAGE Guideline Model : Achievements and Overview. *Journal of the American Medical Informatics Association : JAMIA*, 14(5):589–598.

UICC (2013). How to use the TNM classification. Dernière consultation : Mars 2013. `http://www.uicc.org/sites/default/files/private/How\_to\_use\_TNM\_0.pdf`.

Universität Leipzig (2013). Site Web du Catalogus Professorum Model. Dernière consultation : Mars 2013. `http://www.uni-leipzig.de/unigeschichte/professorenkatalog/`.

Universitatsklinikums Heidelberg (2013). Site Web du projet HELEN. Dernière consultation : Mars 2013. `http://www.klinikum.uni-heidelberg.de/index.php?id=8914`.

URI Planning Interest Group (2001). URIs, URLs, and URNs : Clarifications and Recommendations 1.0. Rapport technique 3305. `http://www.w3.org/TR/uri-clarification/`.

UW Structural Informatics group (2013). Page web de présentation de FMA. Dernière consultation : Mars 2013. `http://sig.biostr.washington.edu/projects/fm/AboutFM.html`.

W3C (2013). Site Web du W3C. Dernière consultation : Mars 2013. `http://www.w3.org/`.

Wang, D., Peleg, M., Tu, S. W., Boxwala, A. A., Greenes, R. A., Patel, V. L. et Shortliffe, E. H. (2002). Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines : : A literature review of guideline representation models. *International Journal of Medical Informatics*, 68(1-3):59 – 70.

Wang, D., Peleg, M., Tu, S. W., Boxwala, A. A., Ogunyemi, O., Zeng, Q., Greenes, R. A., Patel, V. L. et Shortliffe, E. H. (2004). Design and implementation of the GLIF3 guideline execution engine. *Journal of Biomedical Informatics*, 37(5):305 – 318.

Weiss, S., Urso, P. et Molli, P. (2007). Wooki : A P2P wiki-based collaborative writing tool. *In WISE*, pages 503–512.

WHO (2013). Page Web de présentation de CIM. Dernière consultation : Mars 2013. `http://www.who.int/classifications/icd/`.

Wikipatterns.com (2013). Site Web Wikipatterns. Dernière consultation : Mars 2013. `http://www.wikipatterns.com`.

Wikipédia (2013a). Page sur les wikis. Dernière consultation : Mars 2013. `http://fr.wikipedia.org/wiki/Wiki`.

Wikipédia (2013b). Wikipédia, l'encyclopédie libre. Dernière consultation : Mars 2013. `http://fr.wikipedia.org`.

Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M. et Grimshaw, J. (1999). Clinical guidelines : potential benefits, limitations, and harms of clinical guidelines. *BMJ*, 318(7182): 527–530.

Yu, L. (2011). *A Developers Guide to the Semantic Web*. Springer, 1st édition.

zAgile (2013). Page Web de Wikidsmart. Dernière consultation : Mars 2013. `http://www.zagile.com/products/wikidsmart.html`.

Zimmer, A. (2009). Étude et mise en place d'un processus d'édition de connaissances décisionnelles en cancérologie. Rapport de stage de Master SCA, Université Nancy 2.

Résumé

Les connaissances décisionnelles sont un type particulier de connaissances dont le but est de décrire des processus de prise de décision. En cancérologie, ces connaissances sont généralement regroupées dans des guides de bonnes pratiques cliniques. Leur publication est assurée par des organismes médicaux suite à un processus d'édition collaboratif complexe. L'informatisation des guides a conduit à la volonté de formaliser l'ensemble des connaissances contenues de manière à pouvoir alimenter des systèmes d'aide à la décision. Ainsi, leur édition peut être vue comme une problématique d'acquisition des connaissances. Dans ce contexte, le but de cette thèse est de proposer des méthodes et des outils permettant de factoriser l'édition des guides et leur formalisation.

Le premier apport de cette thèse est l'intégration des technologies du Web social et sémantique dans le processus d'édition. La création du wiki sémantique OncoLogiK a permis de mettre en œuvre cette proposition. Ainsi, un retour d'expérience et des méthodes sont présentés pour la migration depuis une solution Web statique. Le deuxième apport consiste à proposer une solution pour exploiter les connaissances décisionnelles présentes dans les guides. Ainsi, le *framework* KCATOS définit un langage d'arbres de décision simple pour lequel une traduction reposant sur les technologies du Web sémantique est développée. KCATOS propose en outre un éditeur d'arbres, permettant l'édition collaborative. Le troisième apport consiste à concilier dans un même système les approches pour la création des guides de bonnes pratiques informatisés : l'approche s'appuyant sur les connaissances symbolisée par KCATOS et l'approche documentaire d'OncoLogiK. Leur fonctionnement conjoint bénéficie des avantages des deux approches.

De nombreuses perspectives sont exposées. La plupart d'entre elles visent à améliorer les services aux utilisateurs et l'expressivité de la base de connaissances. En prenant en compte le travail effectué et les perspectives, un modèle réaliste de système d'aide à la décision complet est proposé.

Mots-clés: Web sémantique, wiki, recommandations de pratique clinique.

Abstract

Decision knowledge is a particular type of knowledge that aims at describing the processes of decision making. In oncology, this knowledge is generally grouped into clinical practice guidelines. The publication of the guidelines is provided by medical organizations as a result of complex collaborative editing processes. The computerization of guides has led to the desire of formalizing the knowledge so as to supply decision-support systems. Thus, editing can be seen as a knowledge acquisition issue. In this context, this thesis aims at proposing methods and tools for factorizing editing guides and their formalization.

The first contribution of this thesis is the integration of social semantic web technologies in the editing process. The creation of the semantic wiki OncoLogiK allows to implement this proposal. Thus, a feedback and methods are presented for the migration from a static web solution. The second contribution consists in a solution to exploit the knowledge present in the decision-making guides. Thus, KCATOS framework defines a simple decision tree language for which a translation based on semantic web technologies is developed. KCATOS also proposes an editor of trees, allowing collaborative editing online. The third contribution is to combine in a single system approaches for the creation of clinical guidelines : the approach based on the knowledge symbolized by KCATOS and the documentary approach symbolized by OncoLogiK. Their joint operation benefits from the advantages of both approaches.

Many future works are proposed. Most of them aim at improving services to users and the expressiveness of the knowledge base. Taking into account the work and prospects, a realistic model to create a decision-support system based on clinical guidelines is proposed.

Keywords: Semantic Web, wiki, clinical practice guidelines.