HAL
open science

K-Ware: vers une gestion conjointe de ressources sémantiques et leurs alignements

Bruno Thiao-Layel, Vianney Jouhet, Gayo Diallo

To cite this version:

Bruno Thiao-Layel, Vianney Jouhet, Gayo Diallo. K-Ware: vers une gestion conjointe de ressources sémantiques et leurs alignements. 6ièmes Journées Francophone sur les Ontologies, Gayo Diallo, Okba Kazar, Fleur Mougin, Oct 2016, Bordeaux, France. hal-01452724

HAL Id: hal-01452724

https://hal.science/hal-01452724v1

Submitted on 2 Feb 2017

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

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

K-Ware: vers une gestion conjointe de ressources sémantiques et leurs alignements

Bruno Thiao-Layel [*], **Vianney Jouhet** [*,**], **Gayo Diallo** [*]

**Univ. Bordeaux, ERIAS - Centre INSERM U1219, F-33000, Bordeaux, France*
***CHU de Bordeaux, Service d'information médicale - Pôle de santé publique, F-33000, Bordeaux, France*
{prenom.nom}@u-bordeaux.fr

RÉSUMÉ. *De plus en plus de sources de données sont annotées au moyen de métadonnées et, de référentiels métiers. C'est le cas dans le domaine biomédical, où l'utilisation de classifications est pregnante (par exemple pour le codage des actes médicaux ou le referencement des analyses biologiques). Ces métadonnées et referentiels qui peuvent être représentés par des ressources termino-ontologiques présentent une certaine hétérogeneité (niveau de formalisme, niveau de granuralité, syntaxe et langue du vocabulaire associé, etc.) qui ne facilite pas la réutilisation secondaire et l'intégration des sources qu'ils annotent. Des approches ont été proposées, souvent sous la forme de hub centralisé, pour offrir des services qui facilitent l'exploration de ces ressources termino-ontologiques en homogénisant leur représentation sous-jacente. Dans ce travail, nous proposons une approche complementaire qui préserve la sémantique des ressources d'origine tout en offrant à l'aide d'un métamodèle générique un support à leur visualisation interactive ainsi que les alignements qui sont générées entre elles. Cette approche, implémentée dans la plateforme K-Ware, est illustrée sur un cas d'usage dans le domaine du cancer.*

ABSTRACT. *There is nowadays a growing number of data sources that are annotated by metadata and standard references. This is the case in the biomedical domain, where the usage of coding systems is usual (for instance for coding medical procedures or referencing biological analysis). These metadata that can be represented by terminologies or ontologies present some heterogeneity (level of formalism, granularity, syntax and language of the associated vocabulary, etc.) which do not facilitate the integration and secondary use of the sources that they annotate. Approaches, often based on a centralized paradigm, have been proposed for providing services which facilitate the exploration of these resources by allowing their homogeneous representation. In this work, we propose a complementary approach which preserves the semantics of the resources being managed while offering through a generic metamodel a support for their interactive visualization together with the alignments that are generated between them. The approach, implemented in the K-Ware platform, is illustrated in the cancer domain.*

MOTS-CLÉS : *ressources termino-ontologiques, métadonnée, alignement, visualisation*

KEYWORDS: *termino-ontological resources, metadata, alignment, visualization*

1. Introduction

Les ressources termino-ontologiques (TOR) sont de plus en plus utilisées comme référentiels et métadonnées dans l'annotation des données des Systèmes d'Information (SI) notamment dans le domaine de la santé. Utiliser de telles ressources permet d'assurer une intéropérabilité sémantique entre différentes sources hétérogènes. Ces TOR peuvent prendre des formes différentes selon l'usage ou encore dans le temps. Il y a un réel enjeu dans l'administration de ces métadonnées et référentiels sémantiques -les TOR- qui représentent des connaissances entre concepts au moyen de relations dans un domaine particulier.

Ces TOR peuvent être représentées dans des formalismes différents selon l'usage et le besoin, allant de simples terminologies à des ontologies très formelles basées sur des langages de représentation comme les logiques de description (Baader *et al.*, 2009).

Plusieurs défis sont à relever concernant ces TOR. Tout d'abord, c'est de pouvoir gérer la persistance de ces TOR en les rendant perennes tout en prenant en compte leur évolution dans le temps ou selon le besoin. Une fois correctement stockées, l'usage de relations hiérarchies provenant, selon le cas, de ressources ontologiques ou terminologiques se doit d'être différencié. En effet, les ressources ont une sémantique d'origine qu'il est nécessaire de préserver. Par exemple, certaines ressources sont structurées autour de la relation de subsomption entre concepts (*est-un* ou *is-a*) tel que spécifié par OWL, alors que d'autres sont structurées autour de relations terminologiques de généricité (*broader*) et de spécificité (*narrower*) tels que spécifiés par SKOS. Il convient dès lors de préserver ces différences lors de l'intégration de TOR hétérogènes. Par ailleurs, gérer des TOR hétérogènes impliquent la gestion de correspondances appelées *alignement* (Shvaiko, Euzenat, 2013) qui améliorent l'intéropérabilité des sources annotées ou décrites par ces TOR. Leur identification et leur manipulation (édition, validation, etc.) parfois complexe représentent un défi. Ensuite, la composition et l'export à la demande de ressources ou parties de ressources sont des services à proposer pour l'interrogation de ces ressources intégrées représentent un autre challenge. La visualisation est un des supports qui permet une exploration interactive dans le cadre de grands graphes que peuvent constituer ces TOR. Enfin, l'enjeu est de fournir des services qui permettent de garder le lien avec les données des sources annotées par ces ressources afin de pouvoir les exploiter au niveau sémantique. Cette mise en lien offre la possibilité d'avoir plusieurs perspectives sur ces sources en changeant uniquement à chaque fois la ressource sémantique utilisée pour l'interrogation.

Bien que ces problématiques se retrouvent dans de nombreuses domaines, celui de la cancérologie en est une illustration tout indiquée dans le cadre de la réutilisation des données déjà rendues disponibles au travers de plusieurs SI hétérogènes. En effet, au cours de la prise en charge d'un patient, de nombreuses informations sont produites par differents acteurs et avec des objectifs differents. Dans un but administratif, le codage des diagnostics pour le Programme français de Médicalisation des Systèmes d'Information (PMSI) s'appuie sur la Classification Internationale des Maladie (CIM10). Par ailleurs, les pathologistes utilisent selon les sites deux terminologies différentes pour la description cyto-pathologique des prélèvement tumoraux: la classification internationnale des maladies pour l'oncologie (CIM-O3) et la codification

des lésions élaborée par l'association pour le Développement de l'informatique en Cytologie et en Anatomie Pathologique (ADICAP). Les sources concernées ne sont donc pas sémantiquement intéropérables du fait de l'utilisation de classifications différentes. Différentes approches ont été abordées pour exploiter ces données (Tognazzo *et al.*, 2009 ; Jouhet *et al.*, 2013) et des travaux ont spécifiquement essayé de mettre en correspondance les différentes terminologies diagnostiques de cancérologie (Brechat *et al.*, 2014). Cependant, ces codes peuvent être utilisés pour caractériser le même diagnostic mais avec un niveau de détail et une structure hétérogène. Il n'y a donc pas réellement de correspondance *parfaite*, mais plutôt *approchante* entre ces codes ce qui nécessite des relations d'appariements complexes. Il faut aussi noter que l'ensemble de ces diagnostics disponibles pour un individu donné est une description ponctuelle d'un processus continu qui se déroule chez le patient. Ainsi l'utilisation de ces données pour suivre l'évolution de la *Maladie Tumorale* d'un individu peut nécessiter la mise en oeuvre de modèles formels permettant de décrire la maladie au regard du diagnostic codé dans les différentes terminologies. Ainsi, afin de permettre l'exploitation de ces données, il est nécessaire de mettre en oeuvre de façon conjointe des terminologies et des modèles formels mais aussi gérer les appariements entre eux.
Notre travail se situe dans cette optique dans le cadre de projets spécifiques où des TOR et des correspondances doivent être intégrées. Cette intégration avec une description fine de chacune des ressources au moyen d'un modèle robuste et évolutif permet l'ajout, la validation, l'exploration et l'exportation à la demande d'entités et de correspondances. L'objectif étant de ne pas être un dépôt centralisé pour des TOR mais un composant qui pourra être intégré dans la mise en oeuvre d'un projet d'intégration de ressources hétérogènes.
La suite de cet article est organisée comme suit. Dans la section 2, nous faisons une brève revue des initiatives récentes dans le domaine. Ensuite dans la section 3 nous présentons les principaux composants et fonctionnalités de K-Ware. Dans la section 4, nous illustrons la mise en oeuvre d'une gestion intégrée de ressources sémantiques, avec des niveaux de formalisme différents, ainsi que leurs alignements dans le domaine de la cancérologie. Enfin nous concluons et traçons quelques perspectives à ce travail.

2. Travaux connexes

Afin d'améliorer la gestion de sources hétérogènes, il faut tout d'abord pouvoir gérer les référentiels et métadonnées les annotant. Les services autour de leur représentation et de leur persistance sont des enjeux à énumérer et à classer selon leur usage, comme l'annotation de sources, la recherche et la validation de correspondances ou encore la capacité à faire évoluer ces référentiels.
Avec l'adoption progressive des technologies du Web Sémantique et la disponibilité croissante d'application à base de connaissances, ces ressources se multiplient et sont de plus en plus larges. Des outils tels que le NCBO BioPortal (Whetzel *et al.*, 2011), Aber-OWL(Hoehndorf *et al.*, 2015), Ontology Lookup Service (OLS) (Jupp *et al.*, 2015) ou encore HeTOP (Grosjean *et al.*, 2011) proposent des services autour d'une

gestion centralisée des TOR pour faciliter leur réutilisation notamment pour l'annotation de données biomédicales, de santé et pour la recherche d'information.
Ces approches reposent sur la même volonté de faciliter l'usage et la sélection de TOR au moyen de services et interfaces permettant l'accès à leurs différentes entités, à leur structure taxonomique mais aussi à leurs métadonnées (comme par exemple les auteurs d'une ressource, le langage de description, l'information sur la version, etc.). Toutefois, des variations sont présentes dans leur manière d'exploiter une ressource, qu'elle soit définie comme une terminologie ou de manière plus formelle avec des axiomes complexes comme dans une ontologie.
Ces variations sont prononcées selon l'implication d'un utilisateur dans les opérations d'import des TOR. En effet, la plateforme ouverte BioPortal, qui a l'objectif d'être un dépôt de TOR dans le domaine biomédical, permet un import aisé et massif de ressources décrites dans les standards du Web Sémantique. Notamment elle permet de traiter des ontologies au format OWL et des terminologies décrites selon le format SKOS. Toutefois, il n'y a alors pas de distinction entre le parcours d'une simple terminologie d'une ontologie (Grosjean *et al*., 2014). En effet, au moment des imports de ressources, BioPortal traite de façon équivalente la relation hiérarchique de SKOS (narrower) et celle de subsomption de OWL. Ainsi l'ensemble des ressources sont considérées comme construites autour de la relation de subsomption quelque soit leur sémantique d'origine. Cette approche peut conduire à des inconsistances. Par exemple dans BioPortal, la classe CIM10 *Sequelae of external causes of morbidity and mortality* est décrite comme *rdfs:subClassOf External causes of morbidity and mortality*. Ce qui n'est pas approprié. La plateforme offre cependant une variété de services pour le dépôt et la réutilisation d'une ontologie persistée.
HeTOP implique des experts pour chacun des imports où est spécifié un modèle propriétaire et où les TOR y sont manuellement maintenues, alignées et traduites. Cet import coûteux humainement procure néanmoins une certaine confiance et une expertise dans les connaissances importées avec des standards du Web Semantique, mais aussi pour d'autres types de données comme des bases de données relationelles. Ce procédé limite l'édition et l'évolution des ressources (Grosjean, 2014). Contrairement à BioPortal, HeTOP ne garde que la dernière version de ses ressources qui plus est, non exportable (Jonquet *et al*., 2016).
Parallèlement, des systèmes et des algorithmes d'alignement de larges ontologies sont mis à disposition [par exemple LogMAP (Jiménez-Ruiz, Grau, 2011), AML (Faria *et al*., 2013), ServOMap (Diallo, 2014), Yam++(Ngo, Bellahsene, 2012)]. Les alignements produits de manière automatisée par de tels systèmes doivent être validés. De même, ils évoluent en fonction de l'évolution des ressources alignées et du contexte d'utilisation. Dès lors, disposer d'infrastructures capables de gérer à la fois et de manière conjointe les ressources termino-ontologiques et leurs alignements devient un enjeu essentiel dans le domaine de l'ingénierie d'ontologies.
Pour l'usage formel des ontologies, on peut citer OLS et Aber-OWL qui ne sont pas uniquement des entrepôts de TOR, mais sont aussi des moyens de parcourir les définitions de concepts au moyen de services reposant sur des raisonneurs. Cela permet non plus uniquement d'explorer des taxonomies et des concepts, mais d'extraire au moyen de leur définition, d'avantage de données alors annotées après inférence de connais-

sances. Ces différentes solutions sont conçues souvent avec une volonté de fournir des hub centralisés pour la mise à disposition de TOR avec des services de visualisation de leur contenu.
Nous proposons alors une approche qui complète ces solutions avec la prise en compte de cette notion de préservation de la sémantique d'origine au moyen de visualisations et d'exploitations des TOR pour permettre le raisonnement poussé sur une ressource quand c'est possible (par exemple une ontologie formelle décrite en OWL) et l'ignorer quand ce n'est pas adapté (par exemple sur une ressource terminologique represéntée en SKOS). L'idée étant de fournir un système facilement instanciable pour des projets spécifiques où il y a un besoin de gestion de ressources sémantiques et leurs alignements.

3. Composants et fonctionnalités de K-Ware

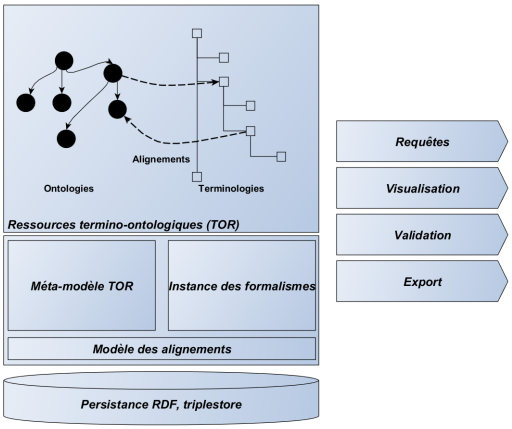

Figure 1. Composants de K-Ware

La conception de K-Ware tient compte de l'existence de TOR décrites avec des niveaux de formalisme différents (exemple de ressources taxonomiques avec des hiérarchies exprimées notamment dans le standard SKOS d'une part, et de ressources exprimées avec des logiques de description comme le OWL). La plateforme se décompose en plusieurs modules où sont gérées au moyen d'un méta-modèle, les TOR et les alignements entre elles. FIGURE 1. Un mécanisme d'héritage des représentations des ressources est introduit à l'image des standards du Web Sémantique où le SKOS est décrit en OWL par exemple. Cela permet la définition simple de nouveaux types de formalismes reposant sur un des types déjà inclus dans K-Ware. Un ensemble de services complète l'infrastructure pour permettre la recherche, la visualisation et l'export de contenu issu de la plateforme ainsi que la validation/édition d'alignements. La persistance des ressources sémantiques et des alignements est assurée à travers un modèle à base de triplets RDF, entreposé dans un triplestore et ainsi interrogeable à l'aide du langage SPARQL.

3.1. Gestion des TOR et leurs alignements

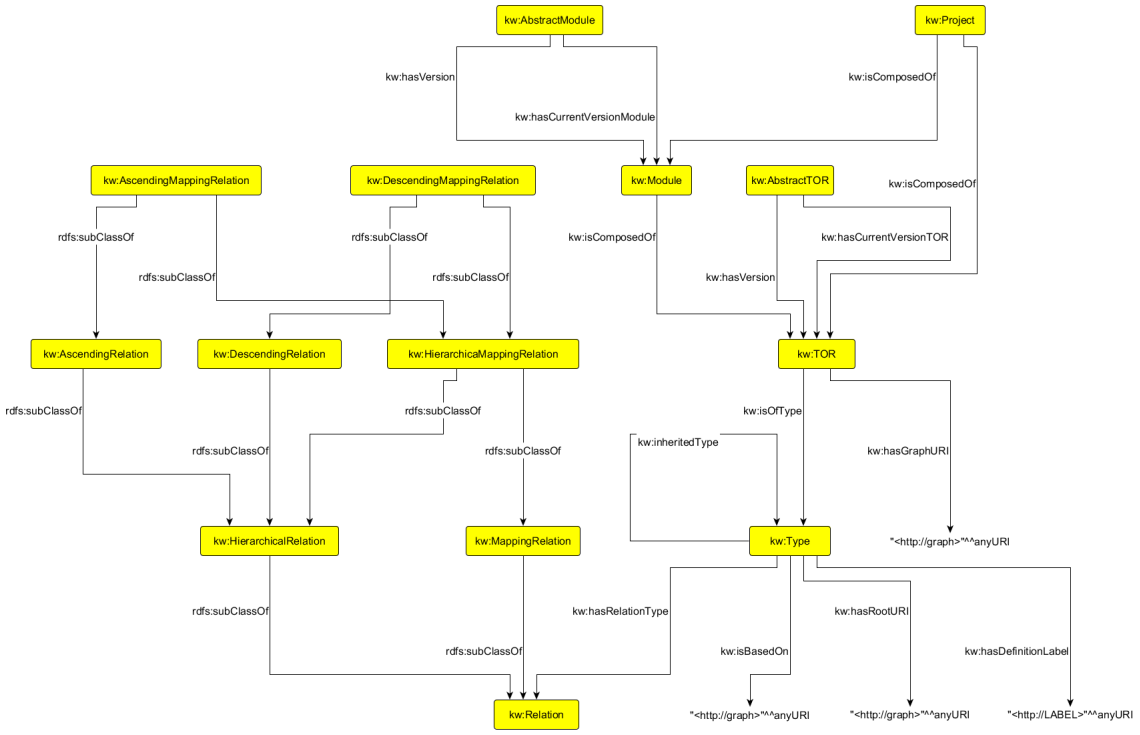

Figure 2. Méta-modèle pour la gestion de TOR et de leurs relations

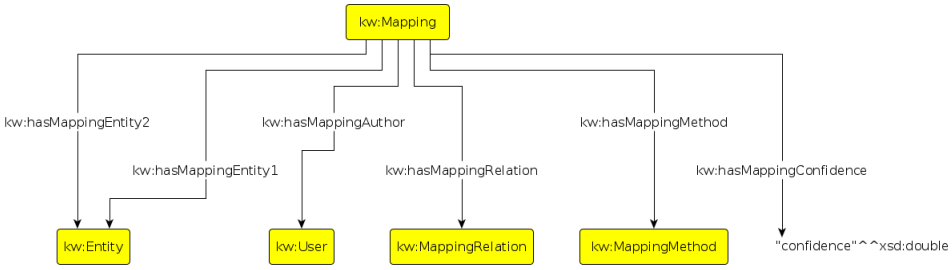

Figure 3. Modèle pour la gestion de tout type d'alignements

Afin de pouvoir gérer de manière adaptée les TOR, tout en préservant leur sémantique d'origine, nous avons défini un méta-modèle qui est instancié pour chaque ressource à intégrer dans la plateforme FIGURE 2. Ainsi, nous avons défini comme entité de haut niveau, la notion de *kw:AbstractTOR*, spécifiant la définition globale et partagée d'une TOR par des versions concrètes manipulables représentées par *kw:TOR*

FIGURE 2. Ce mécanisme permet de caractériser une TOR et prendre en compte par exemple des versions multilingues, ou qui évoluent dans le temps tout en les attachant à un même concept abstrait (Nyulas *et al.*, 2009). Cette notion s'inspire dans son usage à celle de *VirtualOntology* telle que définie dans BioPortal pour la gestion des ontologies. Le méta-modèle formalise la notion de relation hiérarchique, qui peut avoir différente réalisation concrète: la relation de subsomption (is-a), la relation de méronymie (part-of), etc. Cela permet une navigation hiérarchique d'une même ressource selon plusieurs facettes.
Un modèle générique a été défini afin de prendre en compte selon les ressources sémantiques tout type de relations de mappings entre entités (FIGURE 3). Cela permet de gérer par exemple pour les langages OWL et SKOS des relations d'équivalence (e.g *owl:equivalentClass*, *skos:exactMatch*) ou encore des relations d'alignements hiérarchiques (e.g *rdfs:subClassOf*, *skos:broadMatch*).

3.2. Exploration du contenu de la plateforme

L'exploration des ressources persistées est facilitée grâce au méta-modèle instancié et qui permet par la suite de lister les TOR contenus dans la plateforme, de les parcourir en respectant et en utilisant leur sémantique d'origine. En effet, le parcours de chaque TOR est abstrait par les définitions des relations hiérarchiques utilisables pour un *kw:Type*.
Les TOR reposent sur un modèle à base de triplets RDF qui peut être considéré comme une spécification de graphes orientés possédant des arêtes nommées représentant les relations entre des noeuds qui dénotent des classes (concepts). La FIGURE 4 représente l'interface d'exploration de TOR, ici pour la terminologie décrite en SKOS, la *CIMO-03 Morphologie*. L'interface se décompose en trois parties, où on choisit la ressource à afficher avec une relation hiérarchique utilisable, afin de visualiser d'un côté la hiérarchie et de l'autre les mêmes entités mais sous la forme d'un graphe interactif. Ce graphe apporte une visualisation exploratoire, où des interactions utilisateur sont présentes pour l'extension des liens d'un noeud ou encore aider à la sélection d'entités.

4. Illustration sur un cas d'usage en cancérologie

L'utilisation secondaire des données de prise en charge en cancérologie nécessite notamment d'identifier la *Maladie Tumorale* à partir de données hétérogènes, incomplètes et possiblement bruitées. Certaines sources de données en France ont l'avantage d'être structurées (Anatomie et cytologie pathologique, PMSI, réunion de concertation pluridisciplinaire) et l'enregistrement du diagnostic repose sur des terminologies standards. Cependant, différentes spécialités médicales sont amenées à enregistrer ces informations et utilisent des terminologies différentes. La mise en oeuvre de processus automatique d'identification des tumeurs à partir de ces données nécessite donc la gestion de différentes terminologies, de leurs alignements et leur intégration avec un modèle formel permttant de décrire la notion de diagnotic de cancer rapporté à un moment donné par un clinicien et les liens qui existent avec la *Maladie Tumorale* qui

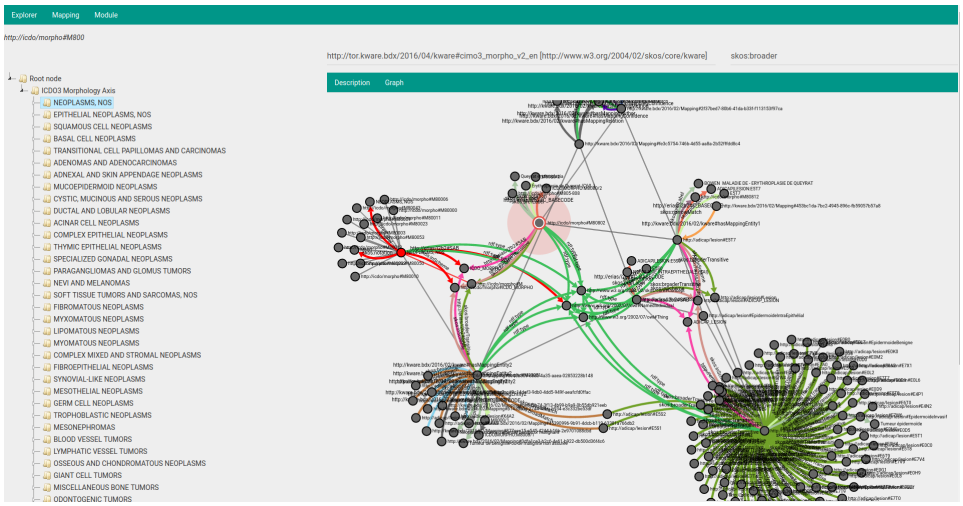

Figure 4. Visualisation du contenu de la terminologie CIM-O3 décrite en SKOS

est un processus continu (FIGURE 5).
On peut alors aborder l'intégration du modèle formel décrivant la *Maladie Tumorale* au moyen de correspondances entre les termes venant des terminologies *ADICAP Lésion et Organe* et *CIMO3 Morphologie et Topographie*. L'exploitation d'une telle intégration se fera au moyen de la sémantique d'origine des ressources à intégrer, en utilisant alors des relations conceptuelles (i.e *rdfs:subClassOf*) pour le parcours de concepts d'une ontologie, et de relations hiérarchiques (i.e *skos:broader* / *skos:narrower*) pour le parcours de termes d'une terminologie, et de la combinaison de ces relations dans son ensemble.
Nous détaillons ci-dessous le processus qui permet d'instancier le cas d'usage tel que présenté en FIGURE 5.

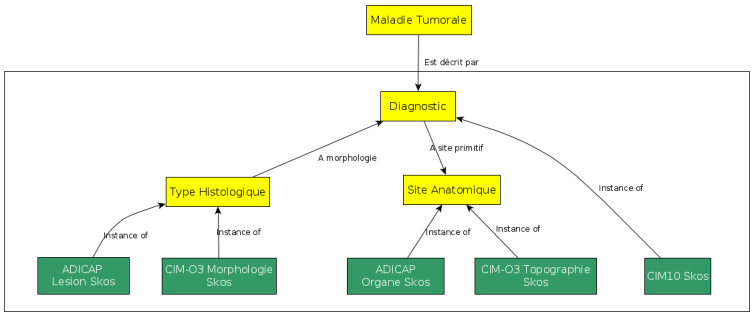

Figure 5. Description de la Maladie Tumorale

4.1. Instanciation dans K-Ware

Nous allons nous intéresser à l'intégration de la partie *morphologique* d'un *Diagnostic* décrivant la *Maladie Tumorale* FIGURE 5. Au moyen du méta-modèle proposé, K-Ware permet l'intégration de ces ressources comme suit. Un utilisateur souhaitant construire le module représentant la combinaison de termes venant de l'*ADICAP Lesion* et de termes venant de la *CIM-O3 Morphologique*, devra au préalable spécifier et caractériser comment ces ressources doivent être utilisées. Dans ce cas d'usage, ces ressources sont représentées par des TOR définies au moyen du vocabulaire SKOS. L'approche retenue dans la FIGURE 6 est alors de n'instancier et prendre en compte que les relations et entités du vocabulaire SKOS utiles et utilisables. Les entités representées en gras et entourés de pointillés sur la figure sont les relations utiles pour le parcours des TOR décrites en SKOS, en prenant en compte toutes les relations hiérarchiques ainsi que les relations d'appariements pouvant aussi être hiérarchiques. La propriété dénotant le terme preféré de chaque entité est aussi pris en compte, ici *skos:prefLabel*. Le mécanisme hiérarchique entre les instances de *kw:Type* est visible

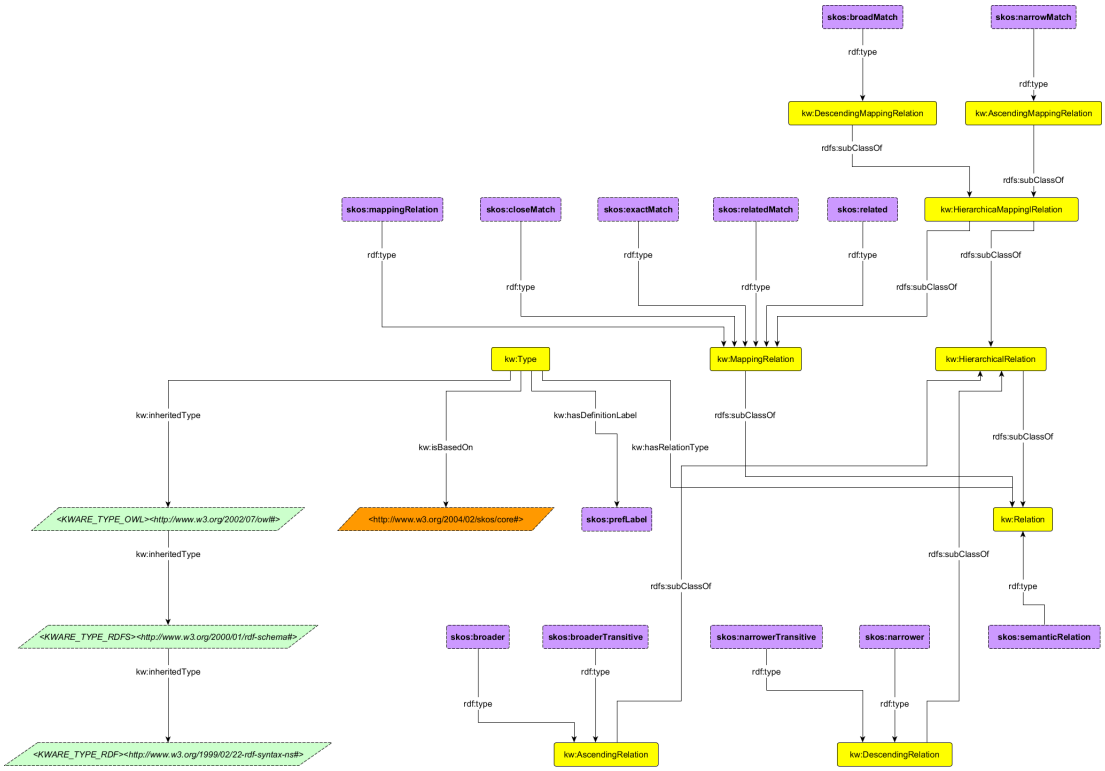

Figure 6. Instanciation du vocabulaire SKOS

avec les entités representées en vert clair et en italique. Chacun des autres vocabulaires est instancié de la même manière.

Les alignements et le modèle représentant *l'intégration morphologique* sont eux gérés de la même manière cependant avec un *kw:Type* pour le vocabulaire OWL. Toute TOR importée se retrouve alors rattachée à la notion abstraite *kw:AbstractTOR* pour la gestion de version. Enfin, chaque composante de cette *intégration morphologique* compose alors l'entité *kw:Module* qui sera composée des (*kw:isComposedOf*) *kw:TOR* concrètes importées et spécifiées avec un type instancié.

4.2. Interrogation

Avec une requête SPARQL FIGURE 7, on peut s'assurer que chaque composante du module *d'intégration morphologique* est spécifiée comme une *kw:TOR* qui ont chacune un graphe nommé (*named graph*) (Carroll *et al.*, 2005) *graphTOR* persisté dans le triplestore et un *kw:Type* définissant leur parcours et leur usage.
Cette encapsulation au travers du méta-modèle permet la composition et la reconnais-

```
PREFIX kw: <http://kware.bdx/2016/04/kware#>

SELECT * WHERE {
  <http://integrationMorphologieDiagnosis_MODULE> kw:isComposedOf ?tor.
  ?tor kw:hasURI ?graphTOR ;
       kw:isOfType ?type .
}
```

kw:Module

kw:TOR	graphTOR	kw:Type
http://erias/MappingLesionMorpho_v1.owl_TOR	http://erias/MappingLesionMorpho_v1.owl	http://www.w3.org/2002/07/owl/kware
http://erias/cimo3_morpho_skos_v2_en.owl_TOR	http://erias/cimo3_morpho_skos_v2_en.owl	http://www.w3.org/2004/02/skos/core/kware
http://erias/modeleCancer_Diagnosis_v6.owl_TOR	http://erias/modeleCancer_Diagnosis_v6.owl	http://www.w3.org/2002/07/owl/kware
http://erias/adicap_lesion_skos_v3_fr.owl_TOR	http://erias/adicap_lesion_skos_v3_fr.owl	http://www.w3.org/2004/02/skos/core/kware

Figure 7. Requête SPARQL montrant la composition du module d'intégration morphologique

sance de plusieurs formalismes différents définissant et représentant les TOR. Différents types d'alignements sont à prendre en compte FIGURE 8. L'abstraction acquise avec le modèle des alignements définie en FIGURE 3 permet de gérer des relations d'équivalence comme *skos:exactMatch* ou encore *skos:broadMatch*.

5. Conclusion

Nous avons brièvement présenté dans cet article la gestion de TOR et de leurs alignements dans le cadre de la plateforme K-Ware en cours de développement et avons illustré une mise en oeuvre dans le cadre de la cancerologie. L'approche permet actuellement à travers la définition d'un méta-modèle de gérer des versions de TOR tout en préservant leur formalisme à l'aide d'un méta-modèle qui abstrait notamment le parcours de polyhiérarchies ainsi que de leurs appariements.

ADICAP Lesion Skos	kw:MappingRelation	CIM-O3 Morphologie Skos
http://adicap/lesion#A0A0	skos:broadMatch	http://icdo/morpho#M8140
http://adicap/lesion#A0A4	skos:broadMatch	http://icdo/morpho#M8550
http://adicap/lesion#A0A5	skos:broadMatch	http://icdo/morpho#M8550
http://adicap/lesion#A0A7	skos:exactMatch	http://icdo/morpho#M8251
http://adicap/lesion#A0B2	skos:exactMatch	http://icdo/morpho#M8149

Figure 8. Extrait des appariements entre des termes ADICAP Lesion et CIM-O3 Morphologie

Comme travaux futurs, nous envisageons dans un premier temps de fournir une interface dediée à la definition fine des types selon la TOR à importer ainsi que l'aide à la validation des entités et de leurs appariements. Par la suite, nous nous focaliserons sur la gestion du couplage connaissances - données annotées. Cela permettra de préparer au niveau sémantique des requêtes à fournir à des sources de données annotés par des connaissances issues de TOR.

Remerciements

Ce travail a reçu un soutien financier de l'agence nationale de sécurité du médicament et des produits de santé (ANSM) dans le cadre du projet Drugs Systematized Assessment in real-life Environment (DRUGS-SAFE).

Bibliographie

Baader F., Horrocks I., Sattler U. (2009). Description logics. In *Handbook on ontologies*, p. 21–43.

Brechat B., Mougin F., Thiessard F., Jouhet V. (2014). Mapping de terminologies diagnostiques en cancérologie par l'intermédiaire du NCI metathesaurus. In *Actes des / proceedings of 15es journées francophones d'informatique médicale co-organisées avec / co-located with 2e congrès national d'informatique médicale (CNIM 2014), fes, morocco, june 12th-13th, 2014.*, p. 34–43.

Carroll J. J., Bizer C., Hayes P., Stickler P. (2005). Named graphs, provenance and trust. In *Proceedings of the 14th international conference on world wide web*, p. 613–622. New York, NY, USA, ACM.

Diallo G. (2014). An effective method of large scale ontology matching. *J. Biomedical Semantics*, vol. 5, p. 44.

Faria D., Pesquita C., Santos E., Palmonari M., Cruz I. F., Couto F. M. (2013). The agreementmakerlight ontology matching system. In *On the move to meaningful internet systems: Otm 2013 conferences*, p. 527–541.

Grosjean J. (2014). *Modélisation, réalisation et évaluation d'un portail multi-terminologique multi-discipline, multi-lingue (3m) dans le cadre de la Plateforme d'Indexation Régionale (PlaIR)*. Thèse de doctorat non publiée, Université de Rouen, Rouen.

Grosjean J., Merabti T., Griffon N., Dahamna B., SJ D. (2011, November). Multiterminology cross-lingual model to create the european health terminology/ontology portal. In *Short papers of the 9th international conference on terminology and artificial intelligence, tia*, p. 119-122. Paris.

Grosjean J., Soualmia L. F., Bouarech K., Jonquet C., Darmoni S. J. (2014). Comparing bioportal and hetop: towards a unique biomedical ontology portal ? In *International work-conference on bioinformatics and biomedical engineering, IWBBIO 2014, granada, spain, april 7-9, 2014.*, p. 227–237.

Hoehndorf R., Slater L., Schofield P. N., Gkoutos G. V. (2015). Aber-owl: a framework for ontology-based data access in biology. *BMC Bioinformatics*, vol. 16, p. 26.

Jiménez-Ruiz E., Grau B. C. (2011). Logmap: Logic-based and scalable ontology matching. In *The semantic web - ISWC 2011 - 10th international semantic web conference, bonn, germany, october 23-27, 2011, proceedings, part I*, p. 273–288.

Jonquet C., Annane A., Bouarech K., Emonet V., Melzi S. (2016). Sifr bioportal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique. In *16th journées francophones d'informatique médicale, jfim'16.*

Jouhet V., Defossez G., Ingrand P. (2013, avril). Automated Selection of Relevant Information for Notification of Incident Cancer Cases within a Multisource Cancer Registry. *Methods of information in medicine*, vol. 52, n° 4.

Jupp S., Burdett T., Leroy C., Parkinson H. E. (2015). A new ontology lookup service at EMBL-EBI. In *Proceedings of the 8th semantic web applications and tools for life sciences international conference, cambridge uk, december 7-10, 2015.*, p. 118–119.

Ngo D., Bellahsene Z. (2012). YAM++ : A multi-strategy based approach for ontology matching task. In *Knowledge engineering and knowledge management - 18th international conference, EKAW 2012, galway city, ireland, october 8-12, 2012. proceedings*, p. 421–425.

Nyulas C., Noy N. F., Dorf M., Griffith N., Musen M. A. (2009). Ontology-Driven Software: What We Learned From Using Ontologies As Infrastructure For Software Or How Does It Taste to Eat Our Own Dogfood. In *Proceedings of the 5th International Workshop on Semantic Web Enabled Software Engineering (SWESE 2009)*, p. 58–72.

Shvaiko P., Euzenat J. (2013). Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, vol. 25, n° 1, p. 158–176.

Tognazzo S., Emanuela B., Rita F. A., Stefano G., Daniele M., Fiorella S. C. *et al*. (2009, février). Probabilistic classifiers and automated cancer registration: An exploratory application. *Journal of Biomedical Informatics*, vol. 42, n° 1, p. 1–10.

Whetzel P. L., Noy N. F., Shah N., Alexander P. R., Dorf M., Fergerson R. W. *et al*. (2011). Bioportal: Ontologies and integrated data resources at the click of a mouse. In *Proceedings of the 2nd international conference on biomedical ontology, buffalo, ny, usa, july 26-30, 2011*.