BAG-GV le graphe de connaissances des bioagresseurs de la vigne

F. Amardeilh[1], S. Bernard[2], R. Bossy[3], C. Corre[4], M. Courtin[3],
X. Delpuech[5], C. Roussey[4], N. Sauvion[6].

[1] Elzeard, Bordeaux, France
[2] LISC, INRAE, Aubière, France
[3] MaIAGE, INRAE, Jouy-en-Josas, France
[4] MISTEA, INRAE, Montpellier, France
[5] IFV, Occitanie, RODILHAN, France
[6] PHIM, Univ Montpellier, INRAE, CIRAD, IRD, Institut Agro, Montpellier, France

florence.amardeilh@elzeard.co, xavier.delpuech@vignevin.com, prenom.nom@inrae.fr

Résumé

Nous avons développé un nouveau graphe de connaissances BAG-GV présentant les bioagresseurs de la vigne à partir de "BioAGgressor Ontology" (BAGO). Ce graphe utilise des éléments de ressources sémantiques existantes dont les classes des taxonomies scientifiques TAXREF-LD et NCBITaxon. Ce nouveau graphe a été développé en utilisant la méthode LOT en partenariat avec 4 experts en viticulture, entomologie et maladie des plantes.

Mots-clés

graphe de connaissances, Web de données, données liées, bioagresseurs des cultures, organismes nuisibles, maladie des plantes, agriculture, bioagresseurs de la vigne, viticulture.

Abstract

We have developed a new knowledge graph, entitled BAG-GV, presenting grapevine bioaggressors based on the 'BioAGgressor Ontology' (BAGO). This graph uses elements from existing semantic resources, including classes from the TAXREF-LD and NCBITaxon scientific taxonomies. This new graph was developed using the LOT method in partnership with four experts in viticulture, entomology and plant diseases.

Keywords

ontology, Web of Data, Linked Open Data, pests, harmful organisms, plant disease, agriculture, grapevine bioaggressors, viticulture.

1 Introduction

L'agronomie et l'agriculture sont confrontées à plusieurs défis sociétaux, économiques et environnementaux majeurs, nécessitant des innovations technologiques. Il existe des sources d'information de référence, contenant des connaissances sur les cultures qui restent difficilement exploitables par les machines : par exemple, des corpus d'observation, comme les bulletins officiels d'alertes agricoles français, appelés *Bulletins de Santé du Végétal* (BSV). Dans ces bulletins, sont entre autres présents, des mentions de maladies, de ravageurs, d'agents pathogènes, de vecteurs [1] , de symptômes visibles et de plantes cultivées, pour informer les lecteurs de l'état sanitaire des cultures. Afin d'annoter les observations des bioagresseurs des cultures et de leurs vecteurs, nous avons besoin d'une ressource spécifique adaptée aux vocabulaires francophones utilisés dans les BSV. Cet article introduit le graphe de connaissances des bioagresseurs de la vigne (BAG-GV) produit à partir de BioAGgressor Ontology (BAGO) [4]. Cette ontologie a été conçue pour modéliser les bioagresseurs des plantes. Ce terme "bioagresseur" fut défini en Août 2018 par le Journal Officiel français comme un "organisme vivant qui cause des dommages aux plantes cultivées ou aux récoltes". Un bioagresseur peut être un organisme nuisible (agent pathogène ou ravageur) ou une plante adventice [2], parfois aussi nommée "ennemi des cultures" dans la littérature. Suivant les principes FAIR, le développement de BAGO et BAG-GV implique une réutilisation stratégique de ressources externes, lorsque cela est possible, notamment des taxonomies scientifiques des organismes vivants comme l'ontologie NCBITaxon [24]. Le processus de développement de BAG-GV a suivi la méthode Linked Open Terms (LOT) et impliqué une collaboration significative avec des experts du domaine. BAGO et BAG-GV sont publiés dans le Linked Open Data (LOD), les rendant ouverts, accessibles pour être réutilisés et interopérables, facilitant ainsi la construction d'applications d'aide à la décision pour la surveillance de la santé des plantes.

1. Un vecteur est soit un organisme vivant (biotique : arthropodes, nématodes, oiseaux, humains, ...) soit un facteur non vivant (abiotique : vent, eau, véhicule de transport,...) qui est capable de transporter avec succès un bioagresseur d'un organisme affecté vers un organisme sain, ou dans l'environnement ou à des aliments de cet organisme sain. Le transport sera considéré comme efficace s'il permet le maintien du bioagresseur dans l'environnement considéré.

2. Une plante adventice est un bioagresseur végétal, qui apparaît dans une parcelle agricole, sans être cultivée. Elle se développe en concurrence avec les plantes cultivées.

ref.	name
[3]	Core Plant Health Threat
[7]	
[16]	Pests in Crops and their Treatments : PCTO[3]
[17]	AgriEnt ontology
[22]	CropPestO ontology[4]
[14]	Rice Disease Ontology : RiceDO[5]
[2]	Plant Disease and Pests Ontology : PDP-O[6]
[6]	EPPO ontology[7]

TABLE 1 – Liste des ontologies de bioagresseurs

ref.	dis.	harm.	vec.	tax.	crop.
[3]	y	ins.	y	UniProt	y
[7]	?	?	?	?	vigne
[16]	n	y	n	ncbi	y
[17]	y	ins.	?	?	y
[22]	y	ins.	n	Agrovoc	verger
[14]	y	y	n	n	riz
[2]	y	y	n	n	dattier
[6]	n	y	n	eppo	y

TABLE 2 – Etude du contenu des ontologies de bioagresseurs

Le reste de l'article est organisé comme suit : La section 2 passe en revue les ressources existantes pour décrire les bioagresseurs de plantes cultivées dont la vigne ; La section 3 illustre la méthode utilisée pour construire le graphe BAG-GV ; La section 4 discute les difficultés rencontrées. La section 5 propose des perspectives.

2 État de l'art sur les ressources décrivant les bioagresseurs de la vigne

L'objectif de BAGO est de proposer une représentation formelle et logique des connaissances sur les bioagresseurs des cultures. Cette représentation sera utilisée pour produire des inférences et faciliter l'intégration de données. C'est la raison pour laquelle notre intérêt concerne les ontologies peuplées qui décrivent les maladies et les bioagresseurs des plantes. Une revue détaillée des ontologies décrivant les bioagresseurs des plantes est disponible dans [5].

Comme illustré dans le tableau 1, nous présentons 8 ontologies peuplées décrivant les bioagresseurs des plantes.

Le tableau 2 indique pour chaque référence sa couverture en terme de : maladies (dis.), organismes nuisibles (harm.), vecteurs (vec.), taxonomie scientifique utilisée (tax.) et culture (crop.). Dans les cellules du tableau, le caractère 'y' signifie 'oui', 'n' signifie 'non', ' ?' sans opinion, 'ins.' veut dire 'insecte nuisible'.

MedISys est une plateforme d'épidémiosurveillance des plantes qui suit l'évolution des ravageurs dans les journaux et les blogs. Ce système utilise l'ontologie 'Core Plant Health Threat' [3]. Malheureusement, cette ontologie n'est pas accessible sur le Web. Nous avons trouvé intéressant le patron qui décrit les symptômes. Il précise l'organe végétal affecté et le type de symptômes. Ces organes sont importés de l'ontologie 'Plant Ontology' (PO) [13]. Les organismes vivants sont décrits par la taxonomie scientifique UniProt [19].

Les travaux de [7] proposent une base de connaissances et des règles d'inférence associées pour la prise de décision en gestion intégrée des ravageurs de la vigne. Les règles implémentent les seuils de la réglementation espagnole. Malheureusement, cette base de connaissances n'est pas accessible sur le Web.

Les travaux de [16] ont produit une ontologie des ravageurs des cultures et de leurs traitements en apprenant des règles à partir de textes rédigés en espagnol. L'ontologie produite est intitulée 'Pests in Crops and their Treatments' (PCTO). PCTO modélise les épidémies comme une relation N-aire entre une culture et un ravageur. PCTO intègre également les traitements de lutte associés. Elle utilise NCBITaxon [24] pour décrire les ravageurs. Elle peut être téléchargée depuis l'article.

AgriEnt est un système d'aide à la décision pour le diagnostic et la lutte contre les insectes ravageurs des principales cultures de l'équateur [18]. Ce système utilise une ontologie "AgriEnt ontology" et un ensemble de règles SWRL [17] pour identifier l'insecte à partir des symptômes observés sur la culture. Cette ontologie n'est pas accessible sur le Web. Elle est basée sur des ontologies précédemment publiées telles que 'Plant Ontology' [13] et IDOplant [27].

Les travaux de [23] développent un système pour aider à identifier les ravageurs des cultures à partir de descriptions textuelles de symptômes rédigées en espagnol. Une fois que l'agriculteur a identifié le ravageur, le système suggère une gamme de traitements. Les cultures concernées sont l'amandier, l'olivier et la vigne. Ce système intègre une ontologie 'CropPestO' [22] et une base de connaissances associée. L'ontologie utilise des concepts du thésaurus Agrovoc pour identifier les cultures et les ravageurs. CropPestO est disponible en téléchargement depuis l'article [23].

Dans [14], les auteurs présentent le développement de l'outil RiceMan : un système expert pour l'identification des maladies du riz et les recommandations de lutte. Le système est basé sur deux ontologies axiomatisées. La 'Rice Disease Ontology' (RiceDO) reconnaît certaines maladies possibles à partir de descriptions précises de symptômes : forme, couleur, partie de la plante, stade de développement de la plante. La 'Treatment Ontology' (TreatO) décrit les traitements en fonction de l'agent causal de la maladie (champignons, bactéries, virus ou ravageurs). La description des symptômes est très précise et pourrait être réutilisée pour d'autres cultures. Ces ontologies sont disponibles sur AgroPortal [15]. Les travaux de [2] présentent un système d'aide à la décision, nommé AgrODSS, pour aider les agriculteurs à diagnostiquer les maladies et les ravageurs, et à prendre des décisions sur les méthodes de lutte appropriées. AgrODSS se concentre uniquement sur la culture du palmier dattier. L'outil de décision est basé sur l'ontologie 'Plant Disease and Pests Ontology' (PDP-O) et son

graphe de connaissances associé. PDP-O est une ontologie OWL axiomatisée avec des classes définies, des propriétés inverses, etc. Ainsi, un raisonneur est appliqué pour inférer de nouveaux faits et enrichir le graphe. Ce graphe est stocké dans GraphDB et est accessible sur le Web [8]. PDP-O et son graphe ont d'abord été construits en analysant et en extrayant des informations d'ontologies et de bases de connaissances existantes : CropPestO [22], Plant Disease Ontology (PDO), RiceDO et TreatO [14] et autres. Le graphe a ensuite été enrichi manuellement en analysant des documents textuels et iconographiques sur les maladies et les ravageurs des cultures et du palmier dattier. PDP-O contient des informations sur les maladies biotiques et abiotiques, leurs agents causaux, les descriptions des symptômes apparaissant sur les parties de la plante, les méthodes de lutte adaptées à une maladie particulière et les preuves qui soutiennent la recommandation des méthodes de lutte. Ainsi, le graphe contient des connaissances non consensuelles et comporte parfois des erreurs. Du fait qu'AgrODSS se concentre uniquement sur la culture du palmier dattier, PDP-O reconstruit la description taxonomique des organismes vivants sans établir de lien avec des ressources telles que la taxonomie NCBI [24]. Notons qu'AgrODSS a été évalué par 11 experts et a obtenu un résultat de 80% de précision globale.

L'entreprise Bayer a transformé la base de données de l'Organisation Européenne et Méditerranéenne pour la Protection des Plantes (OEPP) (mieux connue sous le nom de 'EPPO Global database' [9]) en une ontologie à des fins d'échange interne de données [6]. L'ontologie définit une classe pour chaque ravageur et chaque culture. L'ontologie est disponible en téléchargement sur un dépôt GIT [10].

En résumé, PCTO [16] modélise l'agression d'une plante cultivée par un bioagresseur sous la forme d'une relation N-aire. Les autres ontologies représentent les interactions plante-bioagresseur biologique sous la forme de relations binaires. Chacune de ces ontologies utilise diverses taxonomies scientifiques, mais NCBITaxon [24] est la plus fréquemment utilisée. La 'Plant Disease and Pests Ontology' (PDP-O) et son graphe issu d'AgrODSS [2] fournissent un bon exemple de la nécessité de vérifier, lors du processus de conception, que les connaissances formalisées correspondent à celles des scientifiques, afin d'atteindre une précision maximale. Seule la 'Core Plant Health Threat', utilisée dans la plateforme d'épidémiosurveillance, modélise le vecteur des bioagresseurs. Deux travaux [7], [23] ont permis de construire une base de connaissances des bioagresseurs de la vigne en espagnol.

3 Méthode de création de BAG-GV

Linked Open Terms (LOT) [11] est une méthode utilisée pour le développement d'ontologies et de graphes de connaissances [21]. Cette méthode se concentre sur (1) la réutilisation d'éléments (classes, propriétés et attributs) existants dans des ontologies déjà publiées et (2) la publication de graphes selon les principes du Web de données liées. Elle réutilise trois activités d'ingénierie des connaissances définies dans la méthodologie NeOn [26]. Cette méthode définit les itérations sur les quatre activités suivantes : (1) spécification des besoins, (2) implémentation, (3) publication et (4) maintenance. Une fois l'ontologie BAGO construite, nous avons réitéré les quatre activités pour construire le graphe BAG-GV instanciant BAGO.

3.1 Spécification des besoins de BAG-GV

3.1.1 Questions de compétences

Les besoins ont été spécifiés à l'aide de questions de compétences, illustrées par des exemples de la vigne. Ces questions ont été mises à jour au cours de la phase d'implémentation de l'ontologie BAGO. Dans l'état actuel, BAG-GV répond à 12 questions de compétences :

- Quels sont les taxons scientifiques qui caractérisent l'organisme vivant ? L'organisme vivant est un insecte, un champignon, ...
- Quelle est l'espèce qui caractérise l'organisme vivant ? Ce plant de vigne appartient à l'espèce "Vitis vinifera".
- Quels sont les noms vernaculaires décrivant l'organisme vivant ? Ce plant est une vigne cultivée.
- Quel est le type du bioagresseur ? L'insecte Daktulosphaira vitifoliae" est un ravageur.
- Quel est le rôle d'un organisme vivant impliqué dans une attaque de bioagresseur ? Ce plant de "Vitis vinifera" est l'hôte. L'insecte "Daktulosphaira vitifoliae" est l'agresseur.
- Quelles sont les plantes attaquées par le bioagresseur ? L'insecte "Daktulosphaira vitifoliae" attaque les plants de l'espèce "Vitis vinifera".
- Quels sont les vecteurs qui transportent un agent pathogène ? Par exemple, les cicadelles de l'espèce « Scaphoideus titanus » transmettent le phytoplasme « Candidatus Phytoplasma viti », responsable de la maladie de la flavescence dorée.
- Quel bioagresseur provoque une maladie ? Le mildiou est provoqué par le champignon "Plasmopara viticola".
- Quelles sont les maladies biotiques associées à une plante ? Les maladies biotiques de la vigne sont le phylloxera, le mildiou, etc...
- Quelle est la classe de maladie par type d'agents pathogènes ? Le mildiou est une maladie fongique.
- Quelle partie de la plante présente des symptômes de l'attaque ? Une attaque de phylloxera apparaît sur les racines et les feuilles des vignes.
- Quels sont les symptômes de cette maladie ? Les symptômes du phylloxera sont la présence de galles sur les feuilles et l'apparition de tubérosités sur les racines.

8. https://agrodss.fokus.fraunhofer.de/graphdb/
9. https://gd.eppo.int/
10. https://github.com/basf/EPPOontology
11. https://lot.linkeddata.es/

3.1.2 Cas d'usage de BAG-GV

BAG-GV cherche à représenter les connaissances relatives aux bioagresseurs de la vigne en France ou dans les pays voisins. Il représente les connaissances de l'institut technique IFV et des scientifiques d'INRAE. Ces connaissances doivent être associées à leurs sources d'information. Elles seront publiées sur le Web sous forme de graphe de connaissances afin de respecter les principes FAIR. Ainsi, le graphe servira de données de référence. Il pourra être utilisé pour valider les observations sur le terrain concernant la vigne. Le graphe pourrait être utilisé pour décrire les variables de stress biotique mesurées ou observées dans les expériences sur la vigne.

3.1.3 Identification des sources d'information

Les principales sources d'information ont été le site web Ephytia géré par l'INRAE [12] et les fiches techniques gérées par l'IFV [8]. Afin de nous assurer que nous couvrons la plupart des bioagresseurs de la vigne, nous avons également pris en compte des sources d'information plus techniques telles que :

- le guide phytosanitaire de la viticulture publié par Agroscope [11],
- la deuxième édition de la liste des descripteurs de l'OIV pour les variétés de raisin et les espèces vitis publiée par l'Organisation internationale de la vigne et du vin (OIV) [1],
- la liste interne des principales mauvaises herbes fournie par l'IFV,
- le guide français d'identification des mauvaises herbes en Nouvelle-Aquitaine et de conseils pour leur gestion, intitulé GARANCE, publié par la Chambre d'agriculture de Gironde [10].

Nous avons également repris certaines définitions du glossaire de [25].

BAG-GV présente une vue d'ensemble des bioagresseurs de la vigne apparus en France au cours des dix dernières années. Certaines techniques de NLP ont été appliquées aux bulletins d'alerte agricole français afin d'extraire les bioagresseurs candidats.

3.2 Implémentation de BAG-GV

L'objectif de l'implémentation est de construire le graphe BAG-GV à l'aide du langage formel OWL. Cette activité se déroule en plusieurs sessions qui améliorent progressivement la version du graphe.

3.2.1 Utilisation d'ontologies : BAGO, TAXREF-LD et NCBITaxon

L'une de nos exigences est de réutiliser les classes de TAXREF-LD [20] et NCBITaxon [24], qui stockent toute la taxonomie scientifique. Nous avons fait le choix de décrire tout organisme vivant, utilisé dans BAG-GV, par ses taxons de rang espèce, genre, famille, ordre, classe et règne. Ainsi, nous recherchons dans les ontologies TAXREF-LD et NCBITaxon les classes représentant ces taxons. De plus, la représentation d'une culture doit être associée à son utilisation en réutilisant certains concepts du thésaurus sur l'usage des cultures en France (FCU) [28]. Ainsi, BAG-GV spécialise l'ontologie BAGO pour la vigne.

3.2.2 Conceptualisation

Pour illustrer la conceptualisation, nous présentons les patrons de conception de BAGO instanciés par des exemples de BAG-GV à l'aide du langage CHOWLK [9].

Description d'un organisme vivant La description d'une culture (ou de tout autre organisme vivant), réutilise la classe et les labels issus des ontologies NCBITaxon et TAXREF-LD.

Le taxon le plus spécifique, trouvé dans TAXREF-LD ou NCBITaxon, qui décrit un organisme vivant est représenté par une instance de sa classe taxonomique d'espèce, spécialisation de la classe `bago:GeneticResource`. Par exemple, dans la figure 1, ce taxon, nommé "*bag-gv :taxon_sp_vitis_vinifera*", est décrit par la classe `taxref:taxon/129968` issue de l'ontologie TAXREF-LD et par la classe `ncbi:NCBITAXON/29760` de l'ontologie NCBITaxon.

Cet individu est défini par :

- son label préféré en français, et en anglais, qui contient son nom scientifique de référence ("Vitis vinifera") grâce à la propriété d'annotation `skos:prefLabel`,
- une définition en français, et en anglais, qui contient le rang taxonomique et le nom scientifique ("species taxon Vitis vinifera") à l'aide de la propriété d'annotation `skos:definition`

L'organisme vivant, qui représente la plante cultivée, est représenté par une instance de la classe `bago:LivingOrganism`. Par exemple, dans la figure 1, l'individu nommé "*bag-gv :organism_grapevine*" représente un plant de vigne en général.

Cet individu est défini par :

- son label préféré en français, et en anglais, qui contient son nom scientifique de référence ("Vitis vinifera") grâce à la propriété d'annotation `skos:prefLabel`,
- des labels alternatifs qui contiennent le nom vernaculaire ("vigne") à l'aide de la propriété d'annotation `skos:altLabel`,
- sa définition en français, et en anglais, contenant le nom scientifique ("organisme Vitis vinifera") à l'aide de la propriété d'annotation `skos:definition`.

Cet organisme vivant est décrit par le taxon le plus spécifique. Afin d'ajouter des informations supplémentaires sur les cultures, cet organisme est également décrit par une instance *skos :Concept* issue du thésaurus FCU. Comme le montre la figure 1, l'individu "*bag-gv :organism_grapevine*" est lié à "*bag-gv :taxon_sp_vitis_vinifera*" et "*fcu :Vignes_cultivees*" à l'aide de la propriété d'objet `bago:isDescribedBy`.

Un raisonneur déduit que l'individu "*bag-gv :organism_grapevine*" est typé par la classe `bago:Plant`, grâce à la définition formelle de la

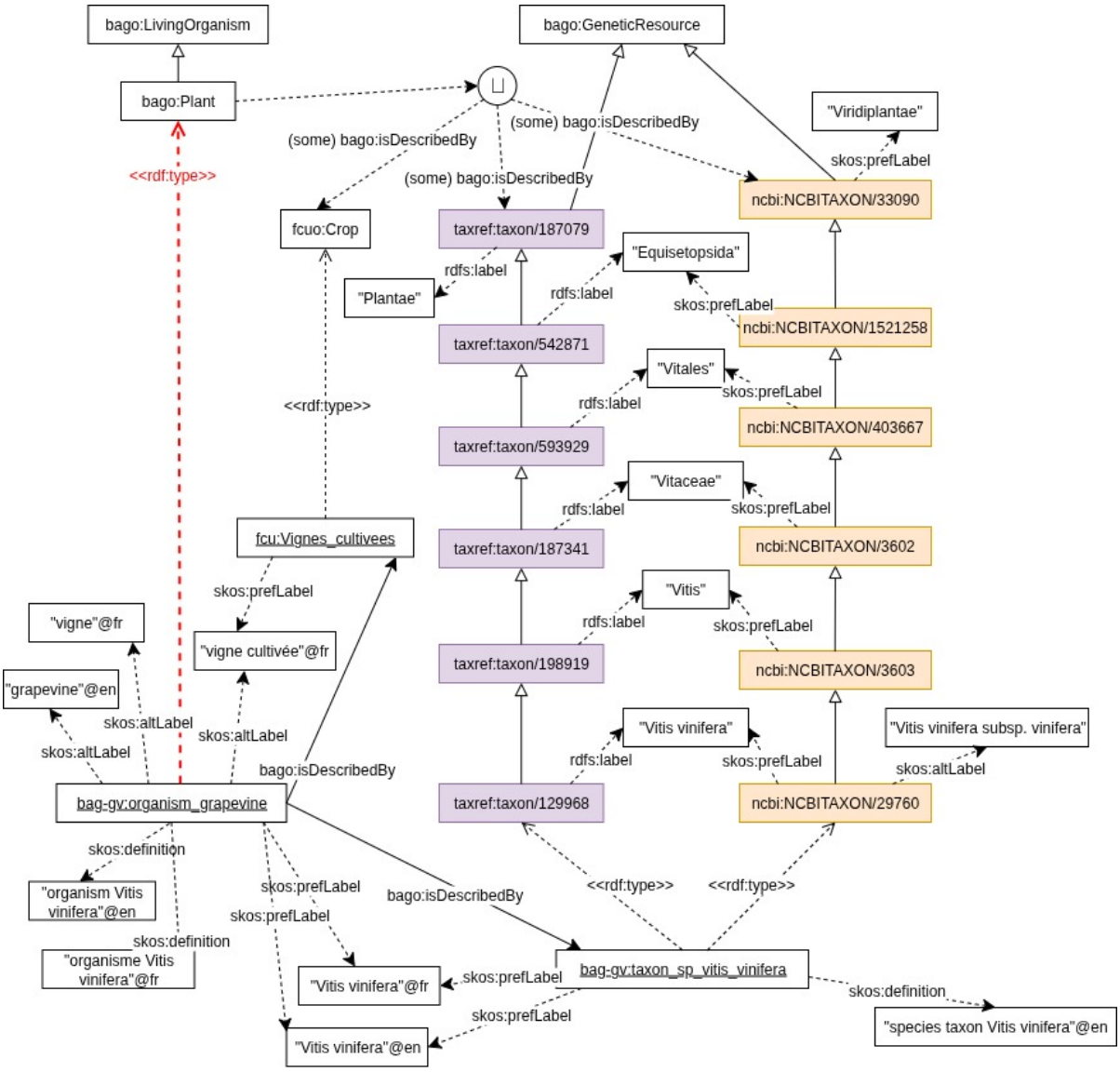

FIGURE 1 – Description d'un plant de vigne cultivée

classe. La figure 1 montre, en rouge, le lien inféré.

Description d'une affection BAGO définit une affection comme "un ensemble d'informations décrivant la situation où plusieurs agents sont en compétition avec une plante. Un agent affecteur nuit au développement d'une plante identifiée (l'affecté) en consommant des ressources nécessaires à la plante et souvent limitées. Des signes de l'affection peuvent être visibles sur la plante affectée (retard de croissance). Les agents responsables de l'affection peuvent être mis en contact avec la plante affectée par le biais d'un ou plusieurs vecteurs." La spécification d'une affection sur une plante cultivée est représentée par un individu, instance de la classe `bago:PlantAffectionSpecification`. Par exemple, dans la figure 2, l'individu nommé "*bag-gv* :$AFGV$006" représente les connaissances que nous avons sur l'affection du liseron des champs sur les vignes.

Cet individu est défini par :

— son label préféré en français, et en anglais, ("affection du liseron des champs sur les vignes") utilisant la propriété d'annotation `skos:prefLabel`,

— sa définition qui contient une description détaillée en français de l'affection précisant la mauvaise herbe, la plante affectée (et ainsi de suite) à l'aide de la propriété d'annotation `skos:definition`,

— un lien vers une ressource Web (voir section 3.1.2) qui décrit la mauvaise herbe à l'aide de la propriété d'annotation `rdfs:seeAlso`.

Tous les organismes vivants qui jouent un rôle dans l'affection sont liés à l'individu « *bag-gv* :$AFGV$006 » :

— l'individu végétal « *bag-gv* :*organism_grapevine* », qui représente la vigne, est lié à « *bag-gv* :$AFGV$006 » à l'aide de la propriété objet `bag:hasAffectedAgent`,

— l'individu végétal « *bag-gv* :*organism_sp_convolvulus_arvensis* », qui représente la plante de liseron des champs, est lié à « *bag-*

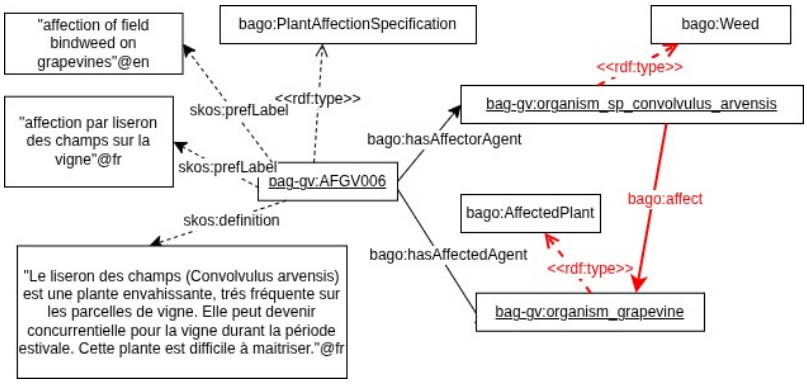

FIGURE 2 – description d'une affection du liseron sur un plant de vigne

gv :$AFGV006$ » à l'aide de la propriété objet `bag:hasAffectorAgent`.

Un raisonneur en déduit que :

— l'individu « *bag-gv* :*organism_grapevine* » est classé dans la classe `bago:AffectedPlant`,

— l'individu « *bag-gv* :*organism_sp_convolvulus_arvensis* » est classé dans la classe `bago:Weed`,

— l'individu « *bag-gv* :*organism_sp_convolvulus_arvensis* » est lié à l'individu « *bag-gv* :*organism_grapevine* » à l'aide de la propriété objet `bago:affect` et de son lien inverse `bago:isAffectedBy` .

La figure 2 montre en rouge les liens déduits.

Description d'une agression BAGO définit une agression comme "Une agression des plantes est un ensemble d'informations décrivant la situation où un ou plusieurs agresseurs attaquent une plante hôte identifiée. L'agression provoque chez la plante hôte une altération de son état ou une modification de son développement. Des signes de l'agression peuvent être visibles sur la plante hôte (les symptômes). Dans certains cas, l'agression peut provoquer une maladie des plantes. Les agresseurs peuvent être mis en contact avec la plante hôte par le biais d'un ou plusieurs vecteurs." La spécification d'une agression par un organisme nuisible sur une plante cultivée est représentée par un individu instance de la classe `bago:PlantAggressionSpecification`. Par exemple, dans la figure 3, l'individu nommé « *bag-gv* :$AGGVD013$ » représente les connaissances dont nous disposons sur l'agression de la « Flavescence dorée » sur les vignes. Cet individu est défini par :

— son label préféré en français, et en anglais, (« agression de la flavescence dorée sur la vigne ») à l'aide de la propriété d'annotation `skos:prefLabel`,

— sa définition contenant une description détaillée en français de l'agression, précisant l'organisme nuisible, la plante hôte, le vecteur (etc.) à l'aide de la propriété d'annotation `skos:definition`,

— une note contenant la description des symptômes en français à l'aide de la propriété d'annotation `skos:note`,

— une note contenant l'organe végétal affecté en français à l'aide de la propriété d'annotation `skos:note`,

— des liens vers une ressource Web décrivant la maladie ou l'organisme nuisible à partir du site Web des fiches techniques Ephytia ou IFV (voir section 3.1.2) à l'aide de la propriété d'annotation `rdfs:seeAlso`.

Une maladie est représentée par un individu, instance de la classe `bago:BioticPlantDisease`. Par exemple, dans la figure 3, l'individu nommé « *bag-gv* :$DGV013$ » représente les connaissances que nous avons sur la maladie de la flavescence dorée. Cet individu est défini par les mêmes informations que l'agression :

— son label préféré en français, et en anglais, (« maladie de la flavescence dorée de la vigne ») à l'aide de la propriété d'annotation `skos:prefLabel`,

— sa définition qui contient une description détaillée en français de l'agression précisant l'organisme nuisible, la plante hôte, le vecteur (etc.) à l'aide de la propriété d'annotation `skos:definition`,

— une note contenant la description des symptômes en français à l'aide de la propriété d'annotation `skos:note`,

— une note contenant l'organe végétal affecté en français à l'aide de la propriété d'annotation `skos:note`,

— des lien vers une ressource Web décrivant la maladie ou l'organisme nuisible à partir du site Web Ephytia ou IFV à l'aide de la propriété d'annotation `rdfs:seeAlso`.

Tous les organismes vivants qui jouent un rôle dans l'agression sont liés à l'individu «*bag-gv* :$AGGVD013$» :

— l'individu de type hôte végétal «*bag-gv* :*organism_grapevine*», qui représente la vigne, est lié à «*bag-gv* :$AGGVD013$» à l'aide de la propriété objet `bag:hasHostAgent`,

— l'individu de type organisme nuisible «*bag-gv* :*organism_sp_candidatus_phytoplasma_vitis*» , qui représente le phytoplasme de la flavescence dorée, est lié à «*bag-gv* :$AGGVD013$» à l'aide de la

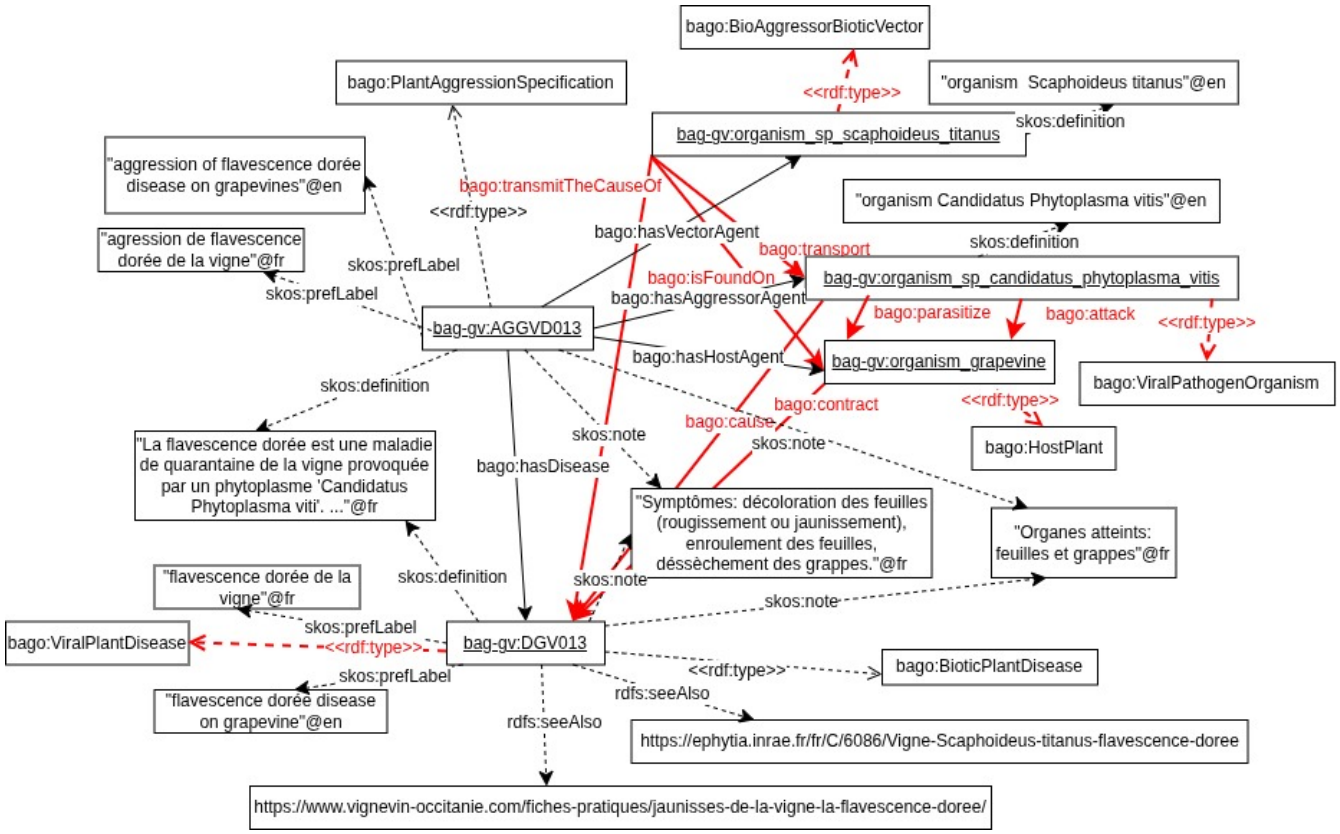

FIGURE 3 – description d'une agression de la flavescence dorée sur un plant de vigne

propriété objet `bag:hasAggressorAgent`,

— l'individu de type organisme vecteur «*bag-gv :organism_sp _scaphoideus_titanus*», qui représente la cicadelle, est lié à «*bag-gv :AGGVD*013» à l'aide de la propriété objet `bag:hasVectorAgent`.

Un raisonneur en déduit que :

— l'individu «*bag-gv :organism_grapevine*» est classé dans la classe `bago:HostPlant`,

— l'individu «*bag-gv :organism_grapevine*» est lié à l'individu «*bag-gv :DGV*013» à l'aide de la propriété objet `bago:contract` et de son lien inverse `bago:isContractedBy`,

— l'individu «*bag-gv :organism_sp_candidatus _phytoplasma_vitis*» est classé dans la classe `bago:ViralPathogenOrganism`,

— l'individu «*bag-gv :organism_sp _candidatus_phytoplasma_vitis*» est lié à l'individu «*bag-gv :DGV*013» à l'aide de la propriété objet `bago:cause` et de son lien inverse `bago:isCausedBy`,

— l'individu «*bag-gv :organism_sp _scaphoideus_titanus*» est classé dans la classe `bago:BioAggressorBioticVector`,

— l'individu «*bag-gv :organism_sp _scaphoideus_titanus*» est lié à l'individu «*bag-gv :organism_sp _candidatus_phytoplasma_vitis*» à l'aide de la propriété objet `bago:transport` et de son lien inverse `bago:isTransportedBy`,

— l'individu «*bag-gv :organism_sp _scaphoideus_titanus*» est lié à l'individu «*bag-gv :organism_grapevine*» à l'aide de la propriété objet `bago:isFoundOn` et de son lien inverse `bago:harbor`,

— l'individu «*bag-gv :organism_sp_scaphoideus _titanus*» est lié à l'individu «*bag-gv :DGV*013» à l'aide de la propriété objet `bago:transmitTheCauseOf` et de son lien inverse `bago:isSpreadBy`,

— l'individu «*bag-gv :DGV*013» est classé dans la classe `bago:ViralPlantDisease`.

La figure 3 montre en rouge les liens déduits.

3.2.3 Encodage de BAG-ORG et BAG-GV

Nous avons remarqué que les descriptions d'organismes vivants présentées dans la figure 1 pouvaient être réutilisées pour produire les graphes de différents bioagresseurs de cultures. Nous avons donc décidé de construire un graphe distinct intitulé BAG-ORG qui stocke la description de tous les organismes vivants et de leurs taxons.

La figure 4 présente le réseau des graphes. BAGO réutilise les classes taxonomiques des règnes ou des domaines issues des ontologies TAXREF-LD et NCBITaxon. BAG-ORG instancie l'ontologie BAGO et réutilise les classes taxonomiques liées aux espèces, genres, familles, ordres et classes issues des ontologies TAXREF-LD et NCBITaxon.

BAG-GV instancie BAGO et importe BAG-ORG.

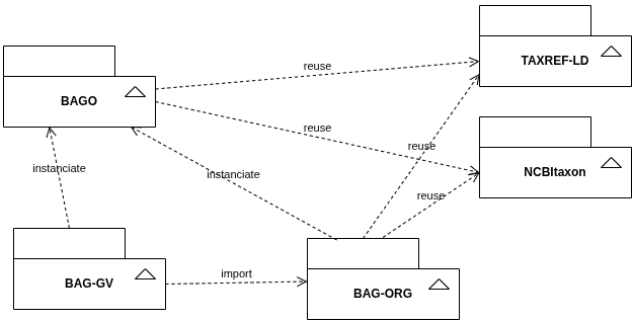

FIGURE 4 – le reseau de graphes

L'objectif est de copier une partie des hiérarchies taxonomiques fournies par des sources externes afin de définir et de décrire les organismes vivants. Ce graphe des organismes vivants, intitulé BAG-ORG, sera partagé avec différents graphes liés aux bioagresseurs spécifiques aux cultures.
Pour encoder nos graphes, nous avons produit manuellement des fichiers tabulés décrivant les organismes vivants, les affections et les agressions. Nous avons utilisé l'outil Protégé avec ses plugins Cellfie [12] et SWRLtab [13] pour générer automatiquement des fichiers RDF/OWL à partir des fichiers tabulés. L'outil Cellfie utilise des règles pour transformer une ligne d'un fichier CSV en triplets RDF. SWRLtab implémente des règles SWRL pour produire de nouveaux triplets à partir d'un graphe existant. Les fichiers RDF/OWL ont ensuite été importés dans un SPARQL endpoint.

3.3 Publication de BAG-ORG et BAG-GV et évaluation

BAG-ORG et BAG-GV sont disponibles sous forme de fichier `.owl` sur le GIT [14]. Chaque graphe dispose de son propre répertoire dans le GIT.
Les graphes sont stockés dans plusieurs SPARQL endpoints :

- L'ontologie BAGO peut être interrogée à l'adresse `https://rdf.codex.cati.inrae.fr/bag-def/sparql`
- Le graphe BAG-GV peut être interrogé à l'adresse `https://rdf.codex.cati.inrae.fr/bag-res/grapevine/sparql`.
- Le graphe BAG-ORG peut être interrogé à l'adresse `https://rdf.codex.cati.inrae.fr/bag-res/organisms/sparql`.

Les SPARQL endpoints sont alimentés par un serveur Apache Jena-Fuseki.
Afin de démontrer que les graphes BAG-ORG, BAG-GV et BAGO répondent aux besoins des utilisateurs, nous avons implémenté les questions de compétence présentées dans la section 3.1 en SPARQL. Toutes les questions de compétence présentées ont pu être traduites en requêtes SPARQL et leurs résultats ont été analysés comme valides par un expert de la vigne. Pour des raisons de concision, l'ensemble des requêtes SPARQL est disponible dans un rapport visible sur le dépôt GIT dans le répertoire grapevine sous le nom "BAG_GV_test_latest.pdf".

3.4 Maintenance de BAG-ORG et BAG-GV

Pour maintenir BAG-GV et BAG-ORG, nous avons besoin des commentaires de la communauté des utilisateurs potentiels. Le dépôt GIT associé permet aux utilisateurs d'écrire des commentaires et de signaler les problèmes rencontrés en déclarant des incidents. Le fichier README fournit les adresses électroniques des gestionnaires et des contributeurs des graphes.

4 Retour d'expériences

Le premier avantage majeur de nos graphes est de réutiliser les hiérarchies existantes de classes représentant les taxons, c'est-à-dire une partie de l'ontologie TAXREF-LD et de celle du NCBITaxon. Ainsi, nos graphes de connaissances sont à jour avec les dernières avancées en matière de taxonomie scientifique. Nous sommes certains que nos graphes de connaissances présentent les noms scientifiques de référence actuels des taxons d'espèces. Nous avons remarqué que la liste des espèces d'adventices fournie par l'IFV contient quelques fautes d'orthographe dans les noms d'espèces que nous avons pu détecter. De plus, nos graphes de connaissances tirent parti des principaux avantages de chaque hiérarchie taxonomique : TAXREF-LD se concentre sur le territoire français et contient les taxons observés sur ce territoire. NCBITaxon est produit par une organisation américaine et est utilisé dans le monde entier. De plus, TAXREF-LD ne contient pas tous les taxons liés au domaine des virus ou des bactéries. NCBITaxon a une couverture plus large et contient tous les taxons que nous recherchons. Ces deux ontologies sont indépendantes et gérées par différents groupes d'experts, ce qui signifie qu'il peut exister certaines différences entre les deux taxonomies. C'est pourquoi, lorsque nous constatons des différences, nous consultons une troisième source d'informations, la base de données mondiale de l'EPPO. C'est le cas pour les noms d'espèces « Cuscuta campestris » et « Cuscuta pentagona ». TAXREF-LD déclare un taxon d'espèce, avec « Cuscuta campestris » comme nom scientifique de référence et « Cuscuta pentagona » comme nom scientifique synonyme. NCBITaxon (et EPPO) déclare deux taxons d'espèces distincts : « Cuscuta campestris » et « Cuscuta pentagona ». Le patron de conception de description des organismes vivants n'impose pas que les deux hiérarchies de classes taxonomiques soient équivalentes. Ainsi, nos graphes de connaissances relient les deux hiérarchies de classes taxonomiques sans créer d'erreurs. La principale difficulté que nous pouvons rencontrer est de rester à jour avec l'évolution et la modification des ontologies externes. Nous devrons donc vérifier que les URI externes vers les-

12. `https://github.com/protegeproject/cellfie-plugin`
13. `https://github.com/protegeproject/swrltab-plugin`
14. `https://forge.inrae.fr/bsv/bio-agressor-ontology`

quelles pointent nos graphes de connaissances existent toujours avec le rang taxonomique correct.
Au cours de la construction de nos graphes BAGO et BAG-GV, nous avons remarqué qu'il existe des différences entre les niveaux d'expertise. Les experts ont du revoir et rediscuter leurs connaissances au cours de la construction des graphes. Par exemple, les pathologistes et les agronomes ne sont pas d'accord sur la définition de symbiose et de parasitisme. Dans nos graphes, nous avons représenté les connaissances des pathologistes de PHIM ayant des définitions plus précises. Un autre exemple est la croyance selon laquelle toutes les maladies proviennent d'agents pathogènes. Il existe un exemple de maladie de la vigne causée par un animal : le phylloxéra. Cet exemple a contraint les pathologistes à revoir la classification des maladies.

5 Synthèse et Conclusion

L'ontologie BAGO représente l'expertise française sur les bioagresseurs des plantes. Elle modélise une agression par une relation N-aire dans le but de pouvoir ajouter de nouvelles informations, comme les conditions climatiques qui favorisent l'apparition d'une maladie.
L'ontologie BAGO a été instanciée et enrichie pour représenter les bioagresseurs de la vigne sous la forme d'un graphe de connaissances intégrant une partie des taxons de TAXREF-LD et NCBITaxon. BAG-GV présente et formalise les connaissances actuelles sur le vignobles français, extraites de diverses sources, mais n'est pas exhaustif. Il aura besoin d'être mis à jour régulièrement, en particulier dans un contexte d'accélération des bioagresseurs émergents (nouveaux ravageurs etc.) en lien avec le changement climatique et la mondialisation des échanges.

Remerciements

Ces travaux ont été financés par le projet CASDAR Connaissances 2024 : Standardiser les données expérimentales et Techniques pour faciliter leur réutilisation et accélérer l'innovation et le développement agricole : application aux travaux sur les biosolutions (STAR)

Références

[1] Adriana Alercia, Rudolf Becher, Jean-Michel Boursiquot, Roberto Carara, Pedro Chomé, Angelo Costacurta, Mirella Giust, Marion Hundemer, Andreas Jung, Thierry Lacombe, Dominique Maigre, Erika Maul, Jesús Ortiz, Anna Schneider, and Andrew Walker. 2nd edition of the oiv descriptor list of grape vine varieties and vitis species, December 2022.

[2] Amani Falah Alharbi, Muhammad Ahtisham Aslam, Khalid Ali Asiry, Naif Radi Aljohani, and Yury Glikman. An ontology-based agriculture decision-support system with an evidence-based explanation model. *Smart Agricultural Technology*, 9 :100659, 2024.

[3] Oscar Alomar, Assumpció Batlle, Josep Maria Brunetti, Roberto García, Rosa Gil, Toni Granollers, Sara Jiménez, Amparo Laviña, Carme Reverté, Jordi Riudavets, et al. Development and testing of the media monitoring tool med is ys for the monitoring, early identification and reporting of existing and emerging plant health threats. *EFSA Supporting Publications*, 13(12) :1118E, 2016.

[4] F Amardeilh, S Bernard, R Bossy, M Courtin, M Hirschy, P Larignon, Catherine Roussey, and N Sauvion. Une ontologie pour modéliser les bioagresseurs des plantes. In Haïfa Zargayouna, editor, *IC2024 : 35es Journées francophones d'Ingénierie des Connaissances*, Plate-Forme Intelligence Artificielle (PFIA), pages 8–13, La Rochelle, France, July 2024. AFIA-Association Française pour l'Intelligence Artificielle, AFIA-Association Française pour l'Intelligence Artificielle.

[5] Florence Amardeilh, Xavier Delpuech, Stéphan Bernard, Robert Bossy, Marine Courtin, Catherine Roussey, and Nicolas Sauvion. Modelisation des interactions plantes organismes vivants sous forme d'ontologie. In *Atelier IA4AgroRun Mobilisation de méthode d'IA pour la sémantisation des données en agroécologie et en veille sanitaire dans le contexte "One Health"*, page 24p., Saint - Denis La Réunion, France, April 2025.

[6] Aarón Ayllón-Benitez, José Antonio Bernabé-Diaz, Paola Espinoza-Arias, Iker Esnaola-Gonzalez, Delphine SA Beeckman, Bonnie McCaig, Kristin Hanzlik, Toon Cools, Carlos Castro Iragorri, and Nicolás Palacios. Eppo ontology : a semantic-driven approach for plant and pest codes representation. *Frontiers in Artificial Intelligence*, 6 :1131667, 2023.

[7] Joaquín Cañadas, Isabel M del Águila, and José Palma. Development of a web tool for action threshold evaluation in table grape pest management. *Precision agriculture*, 18 :974–996, 2017.

[8] IFV Occitanie center. fiches pratiques, 2025. `https://www.vignevin-occitanie.com/fiches-pratiques/` [Accessed on : 2025/04/16)].

[9] Serge Chávez-Feria, Raúl García-Castro, and María Poveda-Villalón. Chowlk : from uml-based ontology conceptualizations to owl. In Paul Groth, Maria-Esther Vidal, Fabian Suchanek, Pedro Szekley, Pavan Kapanipathi, Catia Pesquita, Hala Skaf-Molli, and Minna Tamper, editors, *The Semantic Web*, pages 338–352, Cham, 2022. Springer International Publishing.

[10] Marie Descotis-Bonnaud, Océane Ricau, Violette Aurelle, Alexis Allard, Pauline Burlier, Cyriane Perret, and Adel Bakache. Guide d'aide à la reconnaissance des adventices en nouvelle aquitaine et conseils pour la gestion de l'enherbement (garance). Technical report, Chambre d'griculture de la Gironde, 2021.

[11] Pierre-Henri Dubuis, Aurélie Gfeller, Lina Künzler, Patrik Kehrli, Christian Linder, Jean-Sébastien Reynard, Christophe Debonneville, Jean-Laurent Spring,

Vivian Zufferey, and Kathleen Mackie-Haas. Guide phytosanitaire pour la viticulture 2021–2022. *Agroscope Transfer*, page 57 p., January 2021.

[12] INRAE. Ephytia web site, 2025. `https://ephytia.inrae.fr/fr/CP/30/Cultures-Identifier-connaitre-controler` [Accessed on : 2025/04/16)].

[13] Pankaj Jaiswal, Shulamit Avraham, Katica Ilic, Elizabeth A Kellogg, Susan McCouch, Anuradha Pujar, Leonore Reiser, Seung Y Rhee, Martin M Sachs, Mary Schaeffer, et al. Plant ontology (po) : a controlled vocabulary of plant structures and growth stages. *Comparative and functional genomics*, 6(7-8) :388–397, 2005.

[14] Watanee Jearanaiwongkul, Chutiporn Anutariya, Teeradaj Racharak, and Frederic Andres. An ontology-based expert system for rice disease identification and control recommendation. *Applied Sciences*, 11(21), 2021.

[15] Clément Jonquet, Anne Toulet, Elizabeth Arnaud, Sophie Aubin, Esther Dzalé Yeumo, Vincent Emonet, John Graybeal, Marie-Angélique Laporte, Mark A. Musen, Valeria Pesce, and Pierre Larmande. Agroportal : A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, 144 :126–143, 2018.

[16] Javier Lacasta, F. Javier Lopez-Pellicer, Borja Espejo-García, Javier Nogueras-Iso, and F. Javier Zarazaga-Soria. Agricultural recommendation system for crop protection. *Computers and Electronics in Agriculture*, 152 :82–89, 2018.

[17] Katty Lagos-Ortiz, José Medina-Moreira, César Morán-Castro, Carlos Campuzano, and Rafael Valencia-García. An ontology-based decision support system for insect pest control in crops. In *International Conference on Technologies and Innovation*, pages 3–14. Springer, 2018.

[18] Katty Lagos-Ortiz, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde, José Antonio García-Díaz, and Rafael Valencia-Garcia. Agrient : A knowledge-based web platform for managing insect pests of field crops. *Applied Sciences*, 10(3) :1040, 2020.

[19] Michele Magrane and UniProt Consortium. Uniprot knowledgebase : a hub of integrated protein data. *Database*, 2011 :bar009, 2011.

[20] Franck Michel, Olivier Gargominy, Sandrine Tercerie, and Catherine Faron Zucker. A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF. In *ISWC 2017 Workshop on Semantics for Biodiversity (S4Biodiv 2017)*, volume CEUR Vol. 1933, pages 1–12, Vienna, Austria, October 2017.

[21] María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. Lot : An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, 111 :104755, 2022.

[22] Miguel Ángel Rodríguez-García and Francisco García-Sánchez. Croppesto : An ontology model for identifying and managing plant pests and diseases. In Rafael Valencia-García, Gema Alcaraz-Marmol, Javier Del Cioppo-Morstadt, Néstor Vera-Lucio, and Martha Bucaram-Leverone, editors, *Technologies and Innovation*, pages 18–29, Cham, 2020. Springer International Publishing.

[23] Miguel Ángel Rodríguez-García, Francisco García-Sánchez, and Rafael Valencia-García. Knowledge-based system for crop pests and diseases recognition. *Electronics*, 10(8), 2021.

[24] Eric W Sayers, Jeffrey Beck, Evan E Bolton, Devon Bourexis, James R Brister, Kathi Canese, Donald C Comeau, Kathryn Funk, Sunghwan Kim, William Klimke, et al. Database resources of the national center for biotechnology information. *Nucleic acids research*, 49(D1) :D10, 2021.

[25] Jean-François Soussana. *S'adapter au changement climatique : Agriculture, écosystèmes et territoires*. Éditions Quae, 2013.

[26] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernandez-Lopez. The neon methodology framework : A scenario-based methodology for ontology development. *Applied ontology*, 10(2) :107–145, 2015.

[27] Ramona Walls, Barry Smith, Elser Justin, Goldfain Albert, W Stevenson Dennis, and Pankaj Jaiswal. A plant disease extension of the infectious disease ontology. In *Proceedings of the International Conference on Biomedical Ontology (ICBO-2012)*, pages 1–5, Graz, Austria, 2012.

[28] Nadia Yacoubi Ayadi, Stephan Bernard, Robert Bossy, Marine Courtin, Bill Gates Happi Happi, Pierre Larmande, Franck Michel, Claire Nédellec, Catherine Roussey, and Catherine Faron. A unified approach to publish semantic annotations of agricultural documents as knowledge graphs. *Smart Agricultural Technology*, 8 :100484, 2024.