

Université de Rennes 1

Master 2 Bioinformatique et Génomique

Année universitaire 2016-2017

## **Incorporation des traits relatifs à la santé dans l'ontologie ATOL : enjeu pour les productions animales**

Réalisé par Jérémy Yon

Etablissement d'accueil : INRA, UMR PEGASE, 35590 Saint Gilles

L'étude a bénéficié du soutien financier du GIS Elevage Demain



Encadrants : Marie-Christine Meunier-Salaün

Olivier Dameron

# Remerciements

Je tiens en premier lieu à remercier Marie-Christine Meunier-Salaün qui m'a accueilli pour mon stage de M2.

Je tenais à remercier Olivier Dameron qui a été d'un grand support sur les connaissances du web sémantique.

Je remercie aussi Raphaël Comte, Françoise Thomas, Benoît Peuteman, Steven Mahe, Arnaud Buchet et Anais qui ont partagé leurs pauses et les repas avec moi.

Je remercie aussi Carole Guérin qui m'a accueilli dans son bureau.

## **Table des abréviations**

ATOL : Animal Trait Ontology for Livestock

AHOL : Animal Health Ontology for Livestock

²CSSO: Clinical Terms and Symptom Ontology

NCBI TAXON: National Center for Biotechnology Information Taxonomy

SPARQL: Sparql Protocol and RDF Query Language

OWL: Web Ontology Language

# Table des matières

I.	Introduction .....	1
II.	Matériels et Méthodes.....	3
	1. Matériels.....	3
	1.1 Editeur d'ontologie.....	3
	1.2 Organisation du projet AHOL.....	3
	1.2.1 Structure du projet AHOL.....	3
	1.2.2 Composition des groupes d'experts.....	4
	1.3 Ouvrage de référence.....	4
	2. Méthodes.....	5
	2.1 Workflow du stage.....	5
	2.2 Ontologies connectées.....	7
III.	Résultats.....	7
	1. Séminaire du 2 Janvier à Tours.....	7
	1.1 Présentation de la hiérarchie.....	8
	1.2 Méthode d'échange.....	11
	2. Liste des troubles de santé.....	11
	3 .Détails de la finalité santé.....	12
	3.1 Nombre de nouvelles classes.....	12
	3.2 Nouvelles Propriétés.....	13
	4. Raisonnement sur la finalité santé.....	16
	4.1 Recherche de troubles de santé en fonction de symptômes.....	16
	4.2 Recherche de troubles de santé en fonction d'agents pathogènes.....	17
	4.3 Croisement des deux requêtes précédentes.....	18
IV.	Discussion.....	19
V.	Conclusion.....	22

## Introduction

Au milieu des années 90', on parle pour la première fois du déluge des données <sup>1,2</sup>, les scientifiques de l'époque devant faire face à une déferlante de données provenant du programme sur le génome humain. Les progrès technologiques et les projets ont continué de se succéder en apportant avec eux toujours plus de données. En science de la vie, les données biologiques sont intrinsèquement compliquées et complexes <sup>3</sup>. La complexité de ces données est la résultante de deux facteurs. Tout d'abord, les études biologiques ne se situent pas toutes au même niveau d'étude (séquence, gène, protéine, cellule, organe, organisme et écosystème) ce qui donne des données biologiques très hétérogènes. Ces niveaux sont interdépendants de façon explicite, un gène est codé par une séquence protéique, ou implicite, une relation à un niveau peut nous permettre de mettre en évidence une relation à un niveau voisin <sup>4</sup>. La variabilité est le second facteur de la complexité des données biologiques. Cette variabilité provient de chaque individu ainsi que de ses différents états comme lors d'une pathologie. Aujourd'hui, les données des sciences de la vie sont les plus volumineuses et les plus connectées <sup>5</sup>. En 2016, on dénombre plus de 1500 bases de données biologiques <sup>6</sup>, données qu'il faut rendre réutilisable <sup>7</sup>. Pour faire face aux difficultés de complexité, d'hétérogénéité et de volume il est nécessaire d'intégrer ces données de sources différentes pour permettre un raisonnement sur l'ensemble des données disponibles sur un sujet <sup>8,9</sup>.

Dans le domaine de l'élevage, les informations disponibles dans les bases de données sur les génomes et les performances zootechniques ont explosé. Aussi, il est crucial de disposer d'un langage commun et partagé <sup>10</sup> pour représenter des traits phénotypiques, des symptômes de troubles de santé et autres caractéristiques des animaux de rentes (sexe, stade physiologique, ...). Du fait de l'interdépendance des données biologiques, l'intégration et l'analyse de celles-ci fait appel à des connaissances précises du domaine.

Pour collecter la connaissance des experts du domaine, la solution consiste à annoter les données, générant une standardisation et ainsi mieux faire face à l'hétérogénéité des données des sciences de la vie. Parmi les outils de standardisation à notre disposition, les technologies du Web Sémantique (données liées et ontologies) développées par le World Wide Web Consortium (W3C) apparaissent pertinentes car elles permettent d'intégrer des données de provenances hétérogènes <sup>3</sup>. Parmi les outils, une ontologie est la représentation formelle de la connaissance d'un domaine dans laquelle les termes essentiels sont combinés avec des règles structurantes qui décrivent les relations entre eux <sup>11</sup>. Les ontologies sont destinées à être utilisées par des personnes et des logiciels et permettent la mise en place de

raisonnement automatique. Stocker, partager et réutiliser ses annotations est la clé pour l'analyse systématique des données biologiques. Dans le domaine biomédical, The Open Biomedical Ontology Foundry (OBO)<sup>12</sup> a pour ambition de créer un vocabulaire contrôlé. Un portail<sup>1</sup> rassemblant 562 ontologies (Juin 2017) dont 42 ontologies traitant de traits phénotypiques et 128 ontologies sur la santé a été ainsi créé.

Dans le domaine des productions animales, l'ontologie *Animal Trait Ontology for Livestock (ATOL)*, développé par l'Institut National de Recherche Agronomique (INRA) renseigne sur les caractères phénotypiques des animaux d'élevages (mammifères, oiseaux et poissons). Un caractère phénotypique est la résultante de l'expression du génotype (composition allélique d'un individu) modulée ou non par le milieu environnant. Les traits phénotypiques présents dans l'ontologie ATOL déclinent les caractères d'intérêt pour les finalités majeures de l'élevage : croissance et production de viande, glandes mammaires et production de lait, œuf, foie gras, nutrition, reproduction et bien-être animal. L'ontologie ATOL pour objectif de définir, normaliser, mesurer et référencer avec précision les traits phénotypiques et ainsi servir de référentiel aux programmes de phénotypages des animaux de rentes<sup>13</sup> et améliorer l'analyse des données enregistrées<sup>14</sup>. Jusqu'à présent, les caractères de santé présents dans l'ontologie ATOL sont uniquement considérés au regard de la question du Bien-être animal, négligeant les données relatives aux pathologies et plus largement aux troubles de santé présents dans les élevages. Pour combler ce vide, le projet *Animal Health Ontology for Livestock (AHOL)* financé par le Groupe Interministériel français sur la Sécurité Alimentaire (GISA) a été mis en place sur une durée de un an (démarrage en Janvier 2017) avec l'objectif de concevoir une finalité spécifique à la santé et ainsi enrichir l'ontologie ATOL.

Mon rôle a été triple : 1/ apporter une aide méthodologique aux experts pour structurer leur connaissance sous la forme d'une ontologie, 2/ établir le lien entre la nouvelle finalité et les autres finalités d'ATOL ainsi qu'avec les principales ontologies du domaine et 3/ participer à l'amélioration de l'environnement collaboratif existant. Ce travail mené durant mon stage correspond donc à un travail d'ingénierie plutôt qu'un travail de recherche. Il comportait aussi un rôle de gestion avec l'encadrement des groupes d'experts sollicités pour identifier les caractères phénotypiques, la contribution à l'organisation d'un séminaire et la participation à ce séminaire pour expliquer la démarche suivie dans le projet. Durant ce travail, il a été important aussi d'envisager les prolongations de la construction de la finalité santé

---

<sup>1</sup> <http://bioportal.bioontology.org/>

avec les perspectives de futures analyses qui s'appuieront sur les relations entre les nouveaux termes créés avec ceux qui existaient déjà dans les autres finalités de l'ontologie ATOL.

## Matériels et Méthodes

### 1. Matériels

#### 1.1 Editeur d'ontologie

Le logiciel Protégé<sup>2 15</sup> est un éditeur libre d'ontologie produit par l'université de Stanford et fournissant une interface graphique pour l'utilisateur. Ce logiciel permet la définition des classes, de la hiérarchie des classes et des relations entre les classes et les propriétés de ces relations. Il autorise la personnalisation des formulaires de saisie pour entrer ces données. Le logiciel sauvegarde l'ontologie sous format owl.

#### 1.2 Organisation du projet AHOL

##### 1.2.1 Structure du projet AHOL

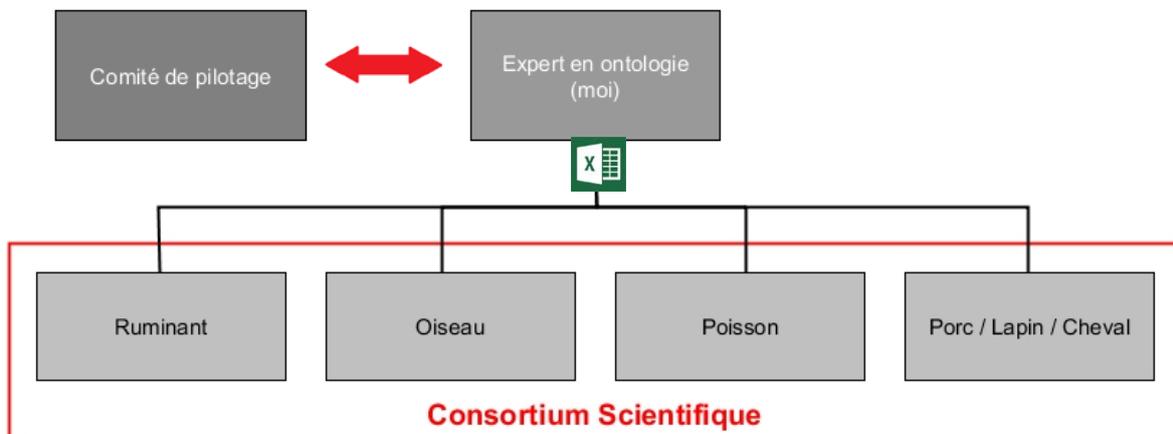


Fig 1. Schéma de la structure du projet AHOL : comité de pilotage, spécialiste ontologie, consortium scientifiques des groupes d'experts.

Le projet AHOL est dirigé par un comité de pilotage constitué de dix personnes provenant de milieux divers (informaticien, vétérinaire, enseignant chercheur, responsable d'installation expérimentale). Ce comité a pour mission de valider les orientations prises durant le projet AHOL. Durant mon stage, l'une de mes tâches a été d'être l'intermédiaire

<sup>2</sup> <http://protege.stanford.edu>

entre ce comité de pilotage et le consortium scientifique composé des experts du domaine de la santé animale.

### *1.2.2 Composition des groupes d'experts*

Pour intégrer la connaissance liée à l'état de santé des animaux et en particulier aux maladies de productions rencontrées en élevage, le travail s'est appuyé sur des personnes possédant la connaissance adéquate pour la création de la finalité santé. Un panel de professionnels de la recherche et de l'enseignement abordant les différentes approches de la santé animale (vétérinaire, enseignant en école vétérinaire et agronomique, chercheur dans les différentes disciplines de la santé, gestionnaire de troupeaux expérimentaux) a donc été contacté. Trente-six d'entre eux sur les 90 personnes sollicitées ont répondu favorablement pour participer au projet AHOL.

Nous avons formé quatre groupes de travail en tenant compte des espèces étudiées par chacun. L'important était d'avoir des personnes travaillant sur la même espèce à des niveaux d'approches différentes afin d'avoir une vision d'ensemble de l'état de santé d'un animal. Les quatre groupes ont travaillé respectivement sur les poissons (G1), les oiseaux (G2), les ruminants intégrant les espèces bovine, caprine et ovine (G3) et enfin les autres espèces regroupant les espèces équine, cunicole et porcine (G4) Au sein de chaque groupe, une personne a été nommée coordinateur pour faciliter les échanges avec le correspondant ontologie, qui était ma fonction. Cette personne est chargé de l'envoi par mail des résultats du groupe. Six personnes du comité de pilotage et possédant des compétences dans le domaine de la santé animale ont intégré les différents groupes en fonction de leur espèce de prédilection dans leurs activités de recherche.

### *1.3 Ouvrage de référence*

En plus de la connaissance des experts, le travail s'est appuyé aussi sur des ouvrages de références. Le Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) a publié en début d'année 2017 un dictionnaire des sciences animales <sup>16</sup>. Cet ouvrage vise à regrouper des informations détaillées sur les sciences animales mais dispersées dans de nombreuses sources. Bien que ce lexique soit présenté sous le nom de dictionnaire, nous sommes en réalité à mi-chemin entre un dictionnaire et une encyclopédie. Ce dictionnaire semi-encyclopédique comporte des domaines utiles pour le projet AHOL, tels que les termes relatifs à la pathologie ou la physiologie, mais aussi de l'information relatives aux autres finalités d'ATOL comme la zootechnie, l'anatomie et l'alimentation. Chaque item

comporte, si possible, le genre du mot, des synonymes, le nom latin, l'étymologie, la traduction anglaise, sa place dans la taxonomie, une définition, le ou les domaines concernés et des indicateurs encyclopédiques (détails et photos). Dans le cadre du projet AHOL qui est un référentiel en anglais, nous avons utilisé cet ouvrage pour proposer une traduction en français et utiliser la définition qui a été traduite en anglais. De plus, le référentiel choisi par ce dictionnaire pour la taxonomie est l'ontologie *NCBI Taxonomy Ontology* que nous utilisons aussi dans la finalité santé de l'ontologie ATOL.

## **2. Méthodes**

### *2.1 Workflow du stage*

Les spécialistes de la santé animale, qui ont participé à la conception de la finalité santé, ont été contactés en fin d'année 2016 pour participer au projet AHOL qui a débuté le 2 Janvier 2017. Le premier échange entre le comité de pilotage et les experts sollicités a été initié lors d'un séminaire à Tours le 2 Février 2017.

Lors du séminaire, les enjeux étaient multiples. Tout d'abord, des présentations ont été préparées visant à initier les experts au principe d'annoter les données afin de mieux les stocker et de les partager à l'aide des ontologies. L'ontologie ATOL et l'intérêt d'y implémenter une finalité santé a été développé par les animateurs du projet. Ma participation a complété ces présentations, et a consisté à décrire le principe de la construction de la finalité santé et la proposition des principales branches de l'ontologie, comme base de travail pour les experts du consortium scientifique. Une répartition des experts dans quatre groupes de travaux a été faite (voir 1.2), suivie d'une discussion visant à valider le périmètre de la nouvelle finalité et la méthode d'échange entre les différents groupes et moi-même. En conclusion de ce séminaire, le périmètre validé incluait les maladies infectieuses et non infectieuses les répandues dans les élevages, ainsi que les agents pathogènes et les symptômes de ces maladies. Il a aussi été demandé aux experts de penser à l'impact que pourrait avoir ces maladies sur les autres traits phénotypiques décrit dans les autres finalités.

Durant la première phase de travail, chaque groupe d'experts devaient sélectionner les troubles de santé à renseigner, tandis qu'en parallèle, j'ai conçu un document excel modèle (appelé « Template », voir Résultats) pour donner un cadre commun à l'ensemble des groupes, facilitant la centralisation des données.

Une réunion par visioconférence pour chaque groupe a été planifiée pour leur présenter le « Template » et commencer à travailler sur la hiérarchie de notre ontologie avec comme question principale : comment ranger au mieux nos maladies ?

La seconde phase de travail avait pour objectif que chaque groupe de travail annote leurs connaissances sur les maladies retenues via le « template ». Pendant ce laps de temps, j'ai finalisé la hiérarchie de la finalité santé et fait la recherche d'ontologies qui traitaient de concepts similaires aux nôtres.

Durant les phases de travail des experts, une communication permanente a été utile entre les groupes et moi-même afin d'assurer une cohérence entre les différents groupes dans leurs annotations. La dernière étape sera de mettre en ligne la nouvelle version de l'ontologie ATOL complétée de la nouvelle finalité santé, ce qui va être fait en prolongement de mon stage.

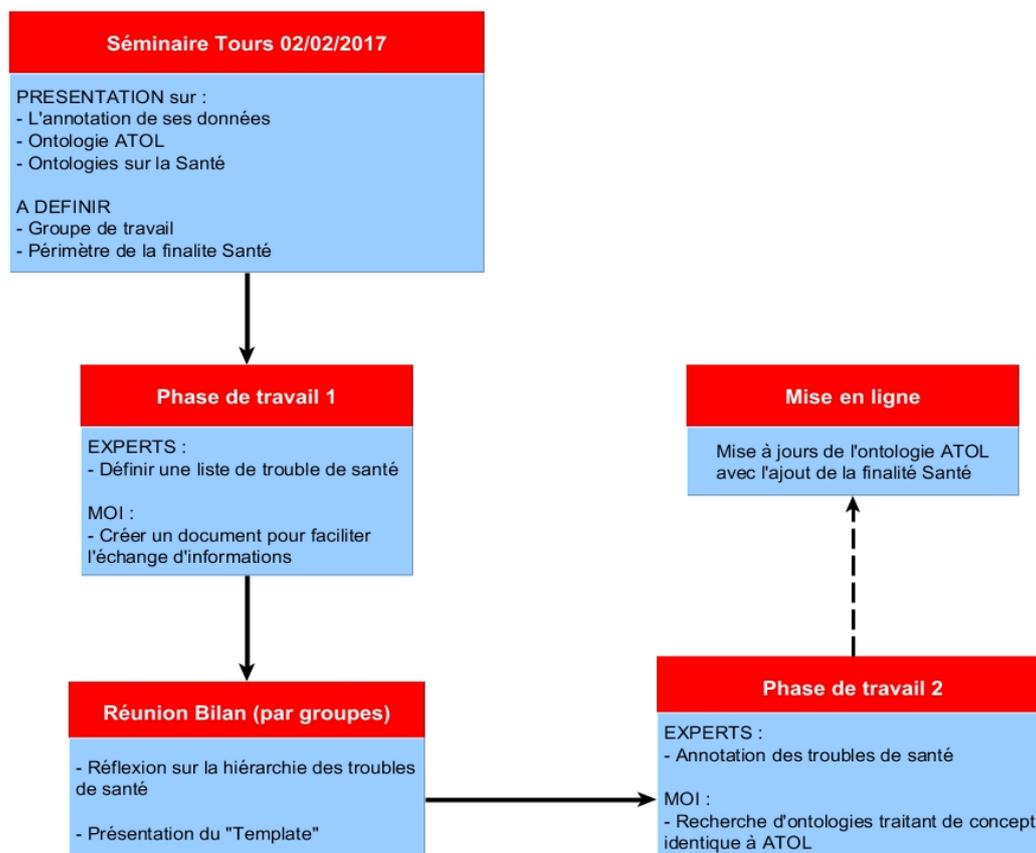


Fig 2. Workflow représentant les principales étapes de mon stage de fin d'étude.

## 2.2 Ontologies connectées

Comme nous vous l'avons présenté dans l'introduction, l'interopérabilité des bases de données est un des enjeux importants pour les sciences de la vie. Afin de respecter cette directive, nous avons interconnecté l'ontologie ATOL à trois autres ontologies. Pour réaliser ceci, j'ai été amené à récupérer les identifiants des ontologies en question et les utiliser au sein de l'ontologie ATOL.

La première ontologie reliée à l'ontologie ATOL est *NCBI Taxonomy Ontology*<sup>17</sup>. Cette ontologie est le référentiel de la classification des espèces vivantes sur terre produit par le Centre Américain pour les Informations Biotechnologique<sup>3</sup>. Nous avons utilisé les identifiants de cette ontologie pour définir les espèces vivantes présentes dans l'ontologie ATOL (animal, bactérie, virus, champignons...).

La seconde ontologie que nous avons décidé d'utiliser est une ontologie regroupant des termes cliniques et des symptômes. Elle se nomme *Clinical Terms and Symptoms Ontology* (CSSO). Les termes renseignés sont classés par anatomie en appareil ou système.

La troisième ontologie est Systematized Nomenclature Of Medicine Clinical Terms (SNOMED CT). C'est un système international de terminologie clinique qui comprend des termes cliniques contrôlés, détaillés et validés médicalement. Pour accéder à SNOMED CT sous format owl, une demande a été faite auprès de l'organisme, ce qui a permis de télécharger le fichier mis à disposition puis exécuter un script Perl, fourni par eux même, et ainsi de créer un fichier owl à partir du fichier d'origine.

## Résultats

### 1. Séminaire 2 Janvier 2017 à Tours

Les objectifs du séminaire étaient de présenter le concept d'ontologie, définir le périmètre de la finalité santé et une méthode d'échange d'informations entre les différents groupes et moi-même. Les troubles de santé observés dans les élevages expérimentaux de l'INRA ont été pris comme point de départ de la réflexion pour le périmètre d'AHOL. Le premier point abordé fut le cadre géographique des troubles de santé rencontrés dans les élevages, en Europe ou ailleurs dans le monde, troubles communs ou rares n'étant pas toujours identiques selon la zone considérée. Aussi nous avons choisi de traiter que les troubles de santé présents sur le territoire européen parce que les experts ont une connaissance étendue de ces troubles chez les espèces de rente, essentiellement à l'échelle européenne.

---

<sup>3</sup> National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>

Ensuite, nous nous sommes intéressés aux informations importantes que nous pouvions relier à un trouble de santé : le caractère infectieux/non infectieux, les symptômes, l'agent pathogène et les différentes modalités des troubles (chronique/aigüe, affecte un stade physiologique/âge particulier de l'animal).

### 1.1 Présentation de la hiérarchie

Les termes présents dans une ontologie (appelé classe) ont été ordonnés de façon hiérarchique. On retrouve trois branches principales pour la finalité santé de l'ontologie ATOL, décrites dans la figure 3 : Health Issue, Organism, Symptom.



Fig 3 Représentation des branches « Health Issue » et « Organism » de la finalité santé de l'ontologie ATOL

La première branche dédiée aux troubles de santé (health issues en anglais) contient des sous-classes qui caractérisent les différences qui existent entre troubles de santé. Les ontologies autorisent l'héritage multiple, de ce fait nous pouvons définir un trouble de santé

comme appartenant à plusieurs sous classes de la branche « troubles de santé ». Lors de la création de ces sous classes, le point important était de les caractériser. Le premier point de différenciation mis en avant par la hiérarchie des troubles de santé est la temporalité du trouble, est-t-il chronique ou aiguë. Le second point concerne la transmissibilité des troubles. Nous avons donc une classe « trouble de santé transmissible » et « trouble de santé non transmissible ». Pour les troubles de santé non transmissibles, nous retrouvons les pathologies métaboliques, psychologiques et les désordres physiques. Pour les troubles de santé transmissibles, nous avons deux sous classes ; les maladies non contagieuses regroupant les maladies vectorielles ou héréditaires et les maladies contagieuses séparées en plusieurs sous classes en fonction de leur type de contagion (direct, indirect ou infectieuse).

La seconde branche regroupe les organismes présents dans l'ontologie ATOL. Dans cette section, on retrouve les agents pathogènes ou vectoriels mais aussi les animaux d'élevage renseignés dans l'ontologie. Dans les autres finalités, la relation entre une classe (ici un trait phénotypique) et un animal de rente est faite via une annotation *Présent* ou *Absent*. Pour la finalité santé, la relation entre une classe (un trouble de santé) et un animal d'élevage est réalisé à l'aide de propriétés décrites dans la partie 3.2. La hiérarchie de cette branche est empruntée à une partie de l'ontologie du NCBI Taxon Ontology. Nous avons seulement récupérer des classes clés qui permettent de différencier tous les organismes présents.

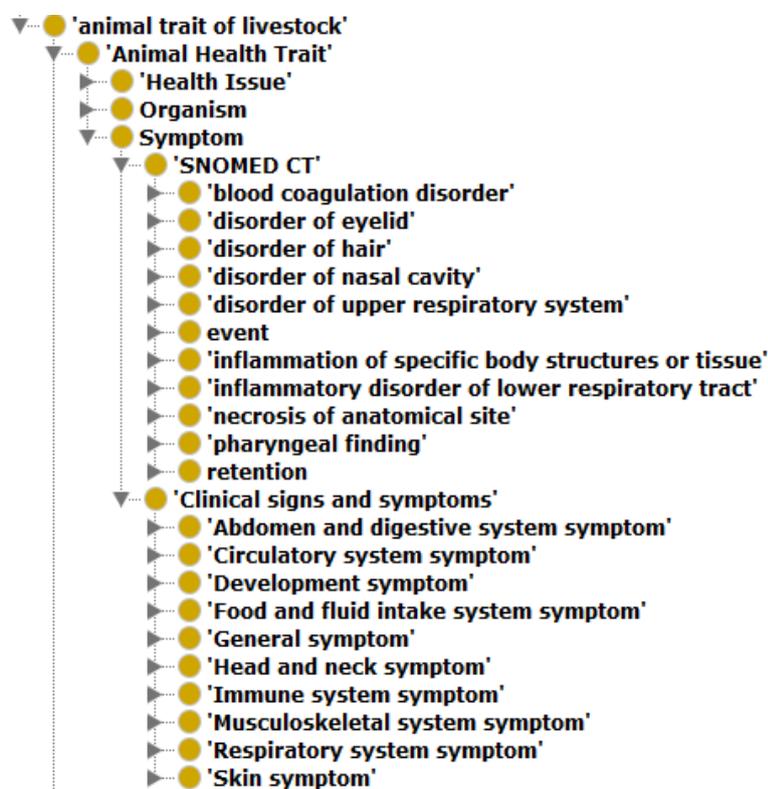


Fig 4 Représentation de branche « Symptom » de la finalité santé de l'ontologie ATOL

La troisième et dernière branche principale de la finalité santé est dédiée aux symptômes des troubles de santé renseignés (Figure 4). Pour une couverture totale des symptômes, nous avons utilisé deux ontologies, qui donnent 24 classes de l'ontologie SNOMED CT et 25 classes de l'ontologie CSSO. Pour reconstruire la hiérarchie à partir de ces ontologies, nous avons utilisé une requête sparql (Figure 5) qui recherche toutes les classes parentes de nos symptômes.

```

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX csso:<http://purl.jp/bio/11/csso/>

CONSTRUCT {
  ?s1 rdfs:subClassOf ?s2 .
  ?s1 rdf:type owl:Class .
  ?s1 rdfs:label ?s1label .
  ?s1 skos:def ?s1def .

  ?s2 rdf:type owl:Class .
  ?s2 rdfs:label ?s2label .
  ?s2 skos:def ?s2def .
}
WHERE {
  ?targetSympt rdfs:subClassOf* ?s1 .
  ?s1 rdf:type owl:Class .
  OPTIONAL { ?s1 rdfs:label ?s1label . }
  OPTIONAL { ?s1 skos:def ?s1def . }

  ?s1 rdfs:subClassOf ?s2 .
  ?s2 rdf:type owl:Class .
  OPTIONAL { ?s2 rdfs:label ?s2label . }
  OPTIONAL { ?s2 skos:def ?s2def . }

  VALUES ?targetSympt {
    csso:CSSO_000064    csso:CSSO_000250    csso:CSSO_000094
    csso:CSSO_000252    csso:CSSO_000075    csso:CSSO_000243    csso:CSSO_000109
    csso:CSSO_000113    csso:CSSO_000115    csso:CSSO_000303    csso:CSSO_000138
    csso:CSSO_000186    csso:CSSO_000187    csso:CSSO_000121    csso:CSSO_000280
    csso:CSSO_000202    csso:CSSO_000206    csso:CSSO_000207    csso:CSSO_000266
    csso:CSSO_000209    csso:CSSO_000244 }
}

```

Fig 5. Requête sparql permettant de reconstruire la hiérarchie de l'ontologie CSSO à partir de symptômes sélectionnés.

Le texte en bleu est la déclaration des Préfix. Les préfix sont les URL des ontologies.  
La partie en orange nous permet d'avoir la réponse de la requête sous cette forme qui correspond au format xml pour un fichier owl.  
La zone verte correspond au corps de notre requête, on y retrouve les contraintes utilisées pour accéder à la réponse souhaitée.  
Le texte en rouge est la déclaration des différentes valeurs que prendra « ?targetSympt ». Les identifiants csso correspondent aux symptômes présents dans l'ontologie ATOL.

### 1.2 Méthode d'échange

Pour faciliter les échanges entre les groupes d'experts et moi-même, j'ai choisi de créer un document Excel qui reprend les informations importantes mises en avant lors du séminaire de Tours (Document Template

- Trouble de santé : nom de la pathologie
- Forme aigüe ou chronique
- Agent pathogène connu
- Modalité de transfert : verticale, horizontale, héréditaire ou vectorielle
- Espèces de l'ontologie ATOL : Porc, Chèvre, Vache, Mouton, Lapin, Cheval, Poule, Canard, Caille, Oie, Dinde, Daurade, Bar, Saumon, Carpe, Truite, Tilapia, Turbot, Poisson Zèbre
- Genre affecté
- Age affecté :
- Stade physiologique affecté :
- Symptôme présent
- Trait de production affecté : classes des autres finalités de l'ontologie ATOL

Fig 6. Catégorie d'informations à compléter dans chacune des colonnes du document Template

## 2. Listes des troubles de santé

Le premier objectif des groupes d'experts a été de définir une liste de troubles de santé qui correspondait au périmètre validé lors du séminaire de Tours. Le groupe « Porc / Lapin / Cheval » a identifié 26 troubles de santé, le groupe « Ruminant » 11 et le groupe « Poisson » 14. Le groupe « Oiseau » n'a pu fournir de listes de troubles de santé à ce jour, en raison d'une indisponibilité des experts engagés sur l'organisation du salon de l'agriculture. Sur la figure 7 ci-dessous, vous pouvez suivre l'avancement du remplissage de la finalité santé. Chaque mesure a été prise le premier de chaque mois.



Fig 7. Taux de remplissage de la finalité santé ( vert : étape de travail réalisé, rouge : à finaliser)

### 3. Détails de la finalité santé

#### 3.1 Nombre de nouvelles classes

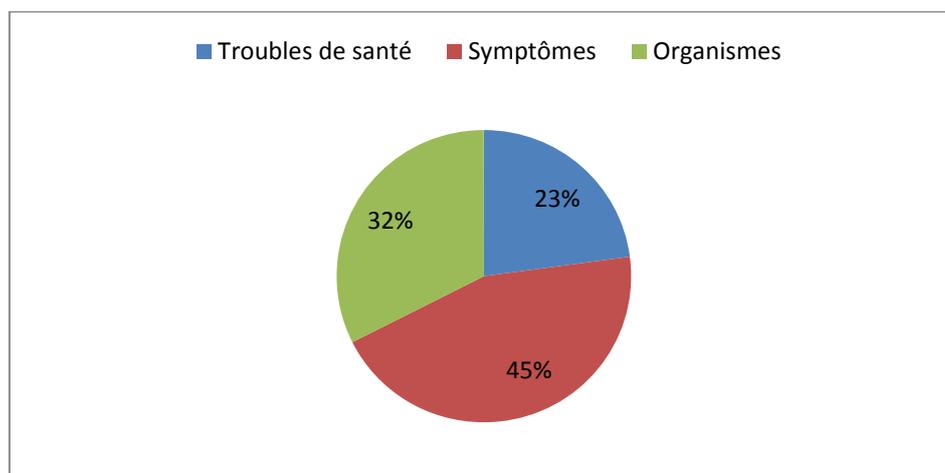


Fig 8. Répartition du nombre de classes par branches principales

La nouvelle finalité de l'ontologie ATOL centrée sur la santé animale est composée de trois branches principales. Le nombre de classes totale de 188 dont 43 pour la branche « trouble de santé », 84 pour la branche « symptôme » et 61 pour la branche « organisme » et une de plus qui est la racine de la finalité santé.

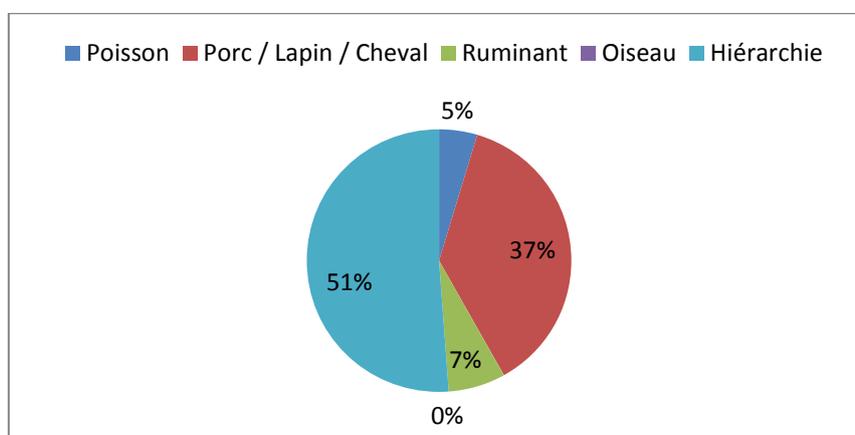


Fig 9. Répartition des classes produites par les groupes d'experts

Au sein de la branche « troubles de santé », les troubles de santé (feuilles de la branche) représente 51% des classes contre 49% pour les classes qui composent la hiérarchie. La répartition des troubles de santé entre les groupes de travail « Ruminant » et « Poisson » est homogène. Le groupe d’experts « Porc / Lapin / Cheval » ont renseigné 16 troubles de santé pour un total de 21 classes. Cette différence est due au faite qu’un trouble de santé peut être représenté par plusieurs classes distinctes suivant son nombre de modalité.

La branche « symptômes » est séparée en deux parties distinctes ; l’une contient les symptômes de l’ontologie SNOMED CT et l’autre contient ceux de CSSO. Au vue de leur démarche différente pour classer leurs symptômes, nous avons dans un premier temps gardé les symptômes rangés sous la hiérarchie pensée par les développeurs de ces ontologies.

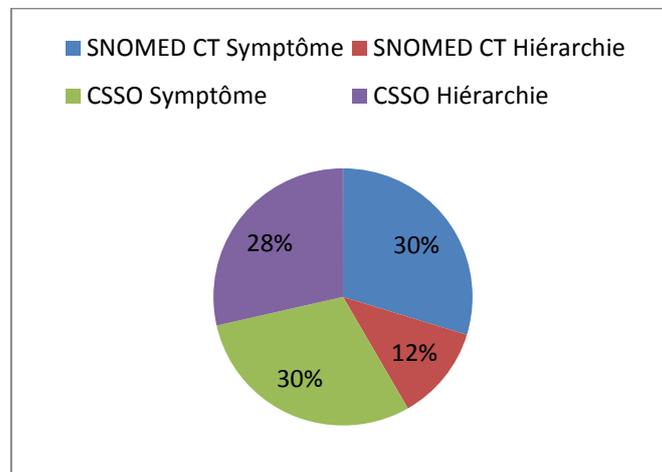


Fig 10. Répartition des symptômes de la branche symptôme

### 3.2 Nouvelles propriétés

Dans le développement d’ATOL, aucune propriété n’avait été développée hormis la propriété « is a », associé à la définition de chaque caractère. Pour la finalité santé, j’ai mis en place onze propriétés, huit propriétés « objets » et trois propriétés « données » (Tableau 1). Les propriétés « objets » sont des relations entre deux classes de l’ontologie alors que les propriétés « données » permettent de créer une contrainte pour une classe. Ces propriétés « données » sont utilisées pour apporter de l’information supplémentaire sur les animaux de rentes (âge, sexe, stade physiologique)

Tableau 1. Tableaux des propriétés incluses dans l'ontologie ATOL

Propriété objet	Domaine	Range
affects	Health issues	Organism
disturbs	Health issues	ATOL finalités*
hasPathogen	Health issues	Organism
hasSymptom	Health issues	Symptom
isAffectedBy	Organism	Health issues
isDisturbBy	ATOL finalités*	Health issues
isPathogenOf	Organism	Health issues
isSymptomOf	Symptom	Health issues
Propriété données	Domaine	Range
Age (weeks)	Organism	Nombre décimal
Gender	Organism	Chaîne de caractère
Physiological Stage	Organism	Chaîne de caractère

\*Atol finalités : reproduction trait, egg trait, animal welfare trait, nutrition trait, fatty liver trait, growth and meat production trait et mammary gland and milk production trait

Pour utiliser ces propriétés j'utilise l'éditeur de classe du logiciel Protégé. Par exemple pour ajouter un agent pathogène à un trouble de santé, il faut cliquer sur la classe souhaitée puis sur le bouton de l'éditeur de classe et d'y écrire la relation voulue avec la bonne syntaxe :  
**Propriété\_objet** some **classe** (**Range**)

Tableau 2. Syntaxe pour l'ajout de propriétés « objets » pour le cas de la Colibacillose chez le Lapin

Syntaxe	Propriété_objet some classe (Range)
Exemple	1. hasPathogen some 'escherichia coli'
	2. hasSymptom some (death or diarrhea or 'growth retardation')
	3. disturbs some ('feed intake' or 'average daily gain')

Pour les propriétés « données », on utilise le même éditeur de classe et la syntaxe est similaire, seul le mot clé change : **Propriété\_données value range**. On peut mélanger les deux syntaxes avec des exemples décrits dans le tableau ci-dessous.

Tableau 3. Syntaxe pour l'ajout de propriétés « objets » et « données » pour le cas de la Colibacillose chez le Lapin

Syntaxe	Propriété_données value classe (Range)
Exemple	1. affects some (Oryctolagus cuniculus and (ageMaximum value 2) and (ageMinimum value 0) and (physiologicalStage value “suckling”))
	2. affects some (Oryctolagus cuniculus and (ageMaximum value 2) and (ageMinimum value 0) and (physiologicalStage value “suckling”))
	3. affects some (Oryctolagus cuniculus and (ageMaximum value 7) and (ageMinimum value 4) and (physiologicalStage value “post weaning”))

Comme on peut le voir sur les différents exemples, j'utilise le mot « *et* » pour ajouter une condition supplémentaire pour les propriétés des données. Avec le terme « *et* », les deux conditions doivent être réunies pour être vrai. Pour déclarer des symptômes affectés à un trouble de santé, j'utilise le mot « *ou* » alors qu'en français, on aurait tendance à citer les symptômes en mettant un « *et* » en mot de liaison. Ici, si nous mettons un « *et* », alors il faut que tous les symptômes soient présents pour trouver le trouble de santé lors d'une recherche par symptômes. Or, tous les symptômes d'un trouble de santé ne sont pas forcément présents en même temps. Le terme « *ou* », nous permet de rester fidèles à la réalité des situations rencontrées en élevage et au regard des réponses biologiques des animaux

## 4. Raisonnement sur la finalité santé

### 4.1. Recherche des troubles de santé en fonction de symptômes

Via des requêtes sparql<sup>4</sup>, nous pouvons questionner l'ontologie afin qu'elle nous renvoie la réponse souhaitée.

Dans l'exemple suivant, décrit dans le tableau ci-dessous, nous allons rechercher les pathologies possédant le symptôme *CSSO\_000202* qui correspond à la septicémie.

Tableau 4. Requête sparql pour la recherche de troubles associés au symptôme « Septicémie »

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX atol:<http://opendata.inra.fr/ATOL/>
PREFIX ncbi:<http://www.ncbi.nlm.nih.gov/taxonomy/>
PREFIX csso:<http://purl.jp/bio/11/ccso/>

SELECT DISTINCT ?disease ?diseaseName ?symptom ?symptomName {
    ?disease      rdfs:subClassOf*      atol:ATOL_0002326 .
    ?disease      rdfs:label            ?diseaseName .
    ?disease      rdfs:subClassOf      ?contrainte .
    ?contrainte   owl:onProperty      atol:hasSymptom .
    ?contrainte   owl:someValuesFrom ?symptom .
    ?symptom      rdfs:label            ?symptomName .

values ?symptom {csso:CSSO_000202}
```

Code Couleur :

Violet : déclaration des Préfix

Rouge : variable attendu pour la réponse

Vert : Contrainte à respecter

Bleu : déclaration de la valeur pour la variable *?symptom*

	disease	diseaseName	symptom	symptomName
1	atol:ATOL_0002360	"Staphylococcosis for new born"	csso:CSSO_000202	"Septicemia"
2	atol:ATOL_0002362	"Streptococcosis"	csso:CSSO_000202	"Septicemia"

Fig11. Réponse à la requête du tableau 4.

<sup>4</sup> Langage de questionnement d'une ontologie

Nous avons deux troubles de santé distincts à savoir Staphylococcie du nouveau-né et Streptococcie qui ont la septicémie en tant que symptôme.

### 5.2. Recherche des troubles de santé en fonction des agents pathogènes

A présent, nous allons rechercher les troubles de santé qui ont la bactérie *Staphylocoque Aureus* en tant qu'agent pathogène.

Tableau 5. Requête sparql pour la recherche de trouble associé au pathogène « *Staphylococcus Aureus* »

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX atol:<http://opendata.inra.fr/ATOL/>
PREFIX ncbi:<http://www.ncbi.nlm.nih.gov/taxonomy/>
PREFIX csso:<http://purl.jp/bio/11/ccso/>

SELECT DISTINCT ?disease ?diseaseName ?pathogen ?pathogenName {
    ?disease      rdfs:subClassOf*      atol:ATOL_0002326.
    ?disease      rdfs:label            ?diseaseName .
    ?disease      rdfs:subClassOf      ?contrainte .
    ?contrainte   owl:onProperty      atol:hasPathogen .
    ?contrainte   owl:someValuesFrom ?pathogen .
    ?pathogen     rdfs:label            ?pathogenName .

values ?pathogen {ncbi:1280}}

```

	disease	diseaseName	pathogen	pathogenName
1	atol:ATOL_0002360	"Staphylococcosis for new born"	ncbi:1280	"Staphylococcus Aureus"
2	atol:ATOL_0002361	"Staphylococcosis for adult"	ncbi:1280	"Staphylococcus Aureus"

Fig 12. Réponse à la requête du tableau 5

Nous avons deux troubles correspondant à l'agent pathogène « *Staphylococcus Aureus* », qui sont deux formes de la Staphylococcie.

### 5.3. Croisement des deux requêtes précédentes

Maintenant, nous allons jumeler les deux requêtes précédentes afin de faire une sélection sur un agent pathogène ainsi que sur un symptôme.

Tableau 6. Requête sparql pour la recherche de trouble associé au pathogène « *Staphylococcus Aureus* » et au symptôme « septicémie »

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX atol:<http://opendata.inra.fr/ATOL/>
PREFIX ncbi:<http://www.ncbi.nlm.nih.gov/taxonomy/>
PREFIX csso:<http://purl.jp/bio/11/ccso/>

SELECT DISTINCT ?disease ?diseaseName ?symptom ?symptomName ?pathogen
?pathogenName {
    ?disease rdfs:subClassOf* atol:ATOL_0002326.
    ?disease rdfs:label ?diseaseName .
    ?disease rdfs:subClassOf ?contrainte .
    ?disease rdfs:subClassOf ?contrainte2 .
    ?contrainte owl:onProperty atol:hasSymptom .
    ?contrainte owl:someValuesFrom ?symptom .
    ?symptom rdfs:label ?symptomName .
    ?contrainte2 owl:onProperty atol:hasPathogen .
    ?contrainte2 owl:someValuesFrom ?pathogen .
    ?pathogen rdfs:label ?pathogenName .
}
Values ?symptom{ csso:CSSO_000202} ?pathogen {ncbi:1280}

```

	disease	diseaseName	symptom	symptomName	pathogen	pathogenName
1	atol:ATOL_0002360	"Staphylococcosis for new born"	csso:CSSO_000202	"Septicemia"	ncbi:1280	"Staphylococcus Aureus"

Fig 13. Réponse à la requête du tableau 6

Au vu des réponses aux requêtes sparql des tableaux 4 et 5, nous n'avons qu'un seul trouble de santé correspondant aux contraintes, Staphylococcie du nouveau-né.

## Discussion

Le Consortium World Wide Web (W3C) ont développé des outils du web sémantique (ontologies, données liées) et ont défini la notion « web des données ». Cette notion vise à favoriser la publication de données structurées, accessibles à travers le web et reliées entre elles pour constituer un réseau global d'informations. En 2000, 38 nouveaux papiers traitant des ontologies ont été publiés sur PubMed<sup>5</sup>, 836 en 2010 et 2272 l'année dernière. Pourtant un secteur comme la production animale ne possédait pas de référentiel avant la parution de l'ontologie ATOL. Le projet AHOL vient compléter l'ontologie ATOL avec une finalité santé animale.

Pour ce projet, nous avons besoin d'accéder aux informations de la santé animale. Des ontologies, comme Gene Ontology<sup>18</sup>, récupèrent les informations de bases de données génomiques publiques. La constitution d'un consortium scientifique avec des correspondant du domaine de la santé animale, a permis d'accéder à une information plus complète et affinée, nécessaire au bon développement de la finalité santé de l'ontologie ATOL.

Dans le domaine de la santé, il existe des ontologies tel que DisGeNET<sup>14</sup> ou Human Disease Ontology<sup>19,20</sup> qui sont centrées sur l'humain. Il existe des ontologies traitant de la santé animale mais qui ne concerne qu'une seule espèce. Il a donc été pertinent de développer une finalité santé pour combler le manque d'informations autour des troubles de santé affectant l'ensemble des animaux de production. Avec le développement de cette nouvelle finalité, nous intégrons de nouvelles informations aux web des données tel que l'association « trouble de santé »-« agent pathogène », « trouble de santé »-« symptôme » et « trouble de santé »-« trait de production affecté ». De plus, dans l'optique d'améliorer l'ensemble des données liées, nous avons connecté l'ontologie ATOL à d'autres ontologies (SnomedCT, CSCO, NCBI Taxon).

Nous pouvons nous demander si des biais ont pu déformer l'information transmises par les experts du projet AHOL. Un biais a pu subvenir lors du passage du contenu du document « template » à son incorporation dans l'ontologie. Ce biais proviendrait d'erreurs de transcription que j'ai pu faire. Afin de contrôler le contenu de la finalité santé, nous pouvons utiliser un module présent dans le logiciel Protégé qui permet de tracer un graphique qui correspond à toutes les relations existantes entre une classe et le reste de l'ontologie. Cette vérification est prévue en continuité du stage avec la création de graphiques pour chacun des troubles présents, transmis aux experts afin qu'ils puissent les valider et confirmer qu'aucun

---

<sup>5</sup> <https://www.ncbi.nlm.nih.gov/pubmed/>

biais ne s'est immiscé dans le processus de construction via le transfert de leurs données par mes soins.

Durant mon stage, j'ai apporté des améliorations à l'ontologie ATOL par le développement d'une nouvelle finalité ciblée sur la santé, intégrant son évaluation par les caractères phénotypiques associées et les troubles qui peuvent observés dans les élevages. J'ai développé des propriétés (objets ou données) qui nous ont permis d'explicitier des relations qui existent entre deux classes différentes via les propriétés « objets » ou fournir plus d'informations pour une classe via les propriétés « données ». Ces relations peuvent être utilisées par la personne qui consulte l'ontologie ou bien par une machine lors de requêtes sparql. L'ajout de liens entre l'ontologie ATOL et d'autres ontologies est aussi une amélioration de l'ontologie existante. Tout comme les relations basées sur les propriétés, ces liens entre ontologies sont exploitables par les hommes et les machines. De plus, ce lien crée une véritable passerelle entre les deux ontologies connectées car il supporte aussi les requêtes sparql. De ce fait, lors de requêtes sparql, nous pouvons utiliser l'intégralité de l'ontologie connectée (classes, propriétés, connexion) pour le questionnement.

Pour continuer de développer un référentiel performant, nous avons plusieurs points où nous pouvons faire des améliorations. Dans la hiérarchie de l'ontologie ATOL, la branche des symptômes est divisée en deux parties distinctes. La première contient les symptômes de l'ontologie Snomed CT et la seconde ceux de l'ontologie Cso. Il serait envisageable d'intégrer ces deux parties en une seule en se basant sur le caractère anatomique des symptômes. Les propriétés « objets » développées durant mon stage nous permettent de créer des relations entre différentes classes. Certaines propriétés comme par exemple « hasSymptom » pourrait être améliorées en créant des sous classes de cette propriété afin de préciser si par exemple le symptôme est systématique ou occasionnel. Le dernier point d'amélioration concerne les ontologies connectées à l'ontologie ATOL. Nous pouvons continuer notre recherche afin d'avoir une meilleur connectivité de nos données sur le web.

A ce jour, le travail des experts a permis de renseigner 56% des troubles sélectionnés, ce qui doit être compléter afin de mettre à jour l'ontologie ATOL sur la finalité santé.

Ce stage a été pour moi une expérience très enrichissante car j'ai, évidemment, appris beaucoup mais j'ai également pu me rendre utile et apporter de nouvelles connaissances et de nouveaux points de vue qui ont permis d'améliorer le projet AHOL dans un esprit pluridisciplinaire. Les différentes personnes engagées dans le projet AHOL sont des spécialistes de la santé animale mais aucune d'entre elles sont des spécialistes des ontologies. Ma contribution au projet avec la gestion des groupes a permis ainsi de leur apporter des

connaissances dans ce domaine. J'ai mis en œuvre aussi mes compétences en informatique pour développer les requêtes. L'ensemble du travail s'est fait en autonomie sur ces questions purement techniques, du fait de la limite des compétences en ontologie et en informatique chez les experts de chaque groupes.

N'étant moi-même pas un spécialiste de la santé animale mais ayant obtenu une licence de biochimie et nutraceutique, j'ai pu appréhender les termes spécifiques du domaine (symptômes, agents pathogènes etc...) grâce notamment à un cours sur l'immunologie, la bactériologie et la virologie suivi au cours de cette licence. Cela m'a permis de bien saisir l'importance des enjeux des ontologies pour les productions animales et de bien comprendre les objectifs du projet AHOL. De plus, le fait d'avoir des connaissances dans ces domaines sans en être spécialiste pour autant, a été un réel avantage pour le projet car mes réflexions étaient moins conditionnées par l'expérience et ont permis d'apporter un regard nouveau sur différentes discussions. Cette complémentarité entre mes connaissances en biologie et mes compétences informatiques est à mon sens, ce que j'ai pu apporter de plus intéressant à la réflexion et à la mise en place du projet AHOL.

Si ma présence a été bénéfique pour l'avancée du projet, je peux affirmer que ce stage a surtout été pour moi une réelle chance et un moment très formateur. Durant ma formation universitaire, j'ai pu acquérir des connaissances à propos du web sémantique que j'ai pu mettre en pratique en réalisant, notamment, la finalité « santé » mais pas uniquement. Ces connaissances m'ont aussi été utiles lors de la conceptualisation de notions appliquées à un thème, en l'occurrence à la santé des animaux de production. De plus, à travers la mise en place de propriétés, j'ai également été amené à m'exercer sur la syntaxe spécifique du langage owl ce qui m'a permis de varier d'avantage encore les différents langages informatiques que je pratique. Des outils du web sémantique tel que le logiciel Protégé ou le Portail développé par OBO m'ont aussi servi et j'ai donc pu m'entraîner à les comprendre et à les maîtriser. Enfin, j'ai amélioré ma connaissance des requêtes sparql également abordées lors de mes cours.

Au-delà de la possibilité de mettre en pratique ce que j'avais appris lors de ma formation, ce stage m'a également apporté une première expérience très enrichissante en management puisque j'ai dû gérer un projet, travailler en coopération avec les différentes équipes de recherche, récupérer leurs résultats, prendre en compte leurs attentes et leurs remarques ou encore les relancer lorsqu'il était nécessaire de le faire. C'était un stage très utile de ce point de vue là puisque j'aimerais par la suite me diriger vers une carrière d'ingénieur et que cette expérience m'a permis de me confronter à la réalité du terrain et de ce que peut être le travail avec plusieurs équipes dont les intérêts, les souhaits et les méthodes de travail ne

sont pas toujours les mêmes. Il m'a fallu prendre le temps d'expliquer ma démarche, de la réadapter, de convaincre mes collègues du bien-fondé de mes choix lorsque c'était nécessaire et de m'imposer parfois mais cela a été pour moi un moment particulièrement formateur et utile pour la suite de mon parcours.

## **Conclusion**

Il a été possible de développer une finalité santé pour l'ontologie ATOL en s'appuyant sur la connaissance de trente-six experts de la santé animale. L'ontologie ATOL avec la finalité santé n'est pas encore disponible sur le web mais il est prévu pour les semaines à venir d'ajouter la finalité santé à la version en ligne. Durant le travail réalisé, nous avons respecté les bonnes conduites à tenir éditées par le W3C afin de proposer un outil performant et connecté au hub des données liées. Les améliorations apportées à l'ontologie permettent l'objectif de l'ontologie ATOL à savoir, être un référentiel pour les programmes de phénotypages des animaux de rentes. De plus, la mise en place de requête sparql offre la possibilité d'élargir le champ des applications qui constituent les perspectives du projet AHOL.

## **Résumé**

Les données des sciences de la vie sont complexes et très nombreuses de ce fait aujourd'hui nous rencontrons des problèmes pour intégrer ces données et les interroger. Le consortium World Wide Web a développé des outils du web sémantique (ontologies et données liées) afin de stocker, partager et analyser les données disponibles. Dans le domaine des ontologies appliquées aux sciences animales, l'ontologie ATOL renseigne insuffisamment la santé animale. A partir de la connaissance d'experts du domaine, j'ai développé une branche santé pour l'ontologie ATOL. L'ontologie ATOL comporte 45 troubles de santé présent dans les élevages européens. Ces troubles ont été ordonnés suivant leur temporalité, leur transmissibilité et leur nature infectieuse ou non. En s'appuyant sur des propriétés « objets » permettant de créer des liens sémantiques entre deux classes, nous avons apporté de l'information supplémentaire pour les troubles de santé telle que son agent pathogène, ses symptômes.

Mots clés : Phénotype, Web Sémantique, Santé Animale, Elevage, Données liées

## **Abstract**

Life sciences datas are complex and very numerous that is why there are problems to integrate and to use and query them nowadays. The World Wide Web Consortium developed semantic web tools (ontologies and linked data) to store, share and analyze available datas. In the field of ontologies applied to animal sciences, ATOL ontology provides insufficient information about animal's health. Helped by experts of this question, I developed a health branch for ATOL ontology. The ATOL ontology would contain 45 health issues that are presents in european farms. Those issues are, as of now, classified depending on their temporality, their transmissibility and their infectious (or not) nature. Based on « objects » properties that allow creating semantic links between two classes, we were able to bring more informations about health troubles like their pathogenic agent, symptoms

Key words: Phenotype, Web Semantic, Animal health, Breeding, Linked data

## Bibliographie

1. Aldhous, P. Managing the genome data deluge. *Science* **262**, 502–503 (1993).
2. Williams, N. Europe opens institute to deal with gene data deluge. *Science* **269**, 630–630 (1995).
3. Bodenreider, O. & Stevens, R. Bio-ontologies: current trends and future directions. *Brief. Bioinform.* **7**, 256–274 (2006).
4. Stevens, R., Goble, C. A. & Bechhofer, S. Ontology-based knowledge representation for bioinformatics. *Brief. Bioinform.* **1**, 398–414 (2000).
5. Stephens, Z. D. *et al.* Big Data: Astronomical or Genomical? *PLoS Biol.* **13**, (2015).
6. Rigden, D. J., Fernández-Suárez, X. M. & Galperin, M. Y. The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Res.* **44**, D1–D6 (2016).
7. Goodman, A. *et al.* Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Comput. Biol.* **10**, e1003542 (2014).
8. Blake, J. A. & Bult, C. J. Beyond the data deluge: Data integration and bio-ontologies. *J. Biomed. Inform.* **39**, 314–320 (2006).
9. Gomez-Cabrero, D. *et al.* Data integration in the era of omics: current and future challenges. *BMC Syst. Biol.* **8**, I1 (2014).
10. Hughes, L. M., Bao, J., Hu, Z.-L., Honavar, V. & Reecy, J. M. Animal trait ontology: The importance and usefulness of a unified trait vocabulary for animal species. *J. Anim. Sci.* **86**, 1485–1491 (2008).
11. Bard, J. B. L. & Rhee, S. Y. Ontologies in biology: design, applications and future challenges. *Nat. Rev. Genet.* **5**, 213–222 (2004).
12. Smith, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251 (2007).

13. Hocquette, J.-F. *et al.* Objectives and applications of phenotyping network set-up for livestock. *Anim. Sci. J.* **83**, 517–528 (2012).
14. Queralt-Rosinach, N., Piñero, J., Bravo, À., Sanz, F. & Furlong, L. I. DisGeNET-RDF: Harnessing the Innovative Power of the Semantic Web to Explore the Genetic Basis of Diseases. *Bioinformatics* btw214 (2016). doi:10.1093/bioinformatics/btw214
15. Musen, M. A. The Protégé Project: A Look Back and a Look Forward. *AI Matters* **1**, 4–12 (2015).
16. Meyer. Dictionnaire des Sciences Animales. (2017). Available at: <http://dico-sciences-animales.cirad.fr/>. (Accessed: 16th May 2017)
17. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **40**, D136–D143 (2012).
18. Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucleic Acids Res.* **36**, D440–D444 (2008).
19. Schriml, L. M. *et al.* Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**, D940-946 (2012).
20. Kibbe, W. A. *et al.* Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* **43**, D1071-1078 (2015).

## **Présentation de la structure d'accueil**

L'institut National de la Recherche Agronomique (INRA) a été créé en 1946. L'INRA dispose de 17 centres en France avec 150 sites de recherche et d'expérimentation et est le premier institut de recherche européen pour l'agriculture, l'alimentation et l'environnement.

Mon Stage s'est déroulé sur le site de l'INRA St Gilles dans l'équipe SysPorc qui travaille au développement de systèmes et d'outils contribuant à l'émergence de pratiques innovantes et durables.