

# Y a-t-il une vie pour les données au-delà de nos systèmes d'informations ?

Bernadette Urban<sup>1</sup>, Edmond Ricard<sup>2</sup>

**Résumé.** Cet article décrit sommairement comment et pourquoi les bases d'ontologies concernant les animaux de rente peuvent se révéler être un bon vecteur d'interopérabilité entre les données stockées dans les systèmes d'informations (SI), développés et maintenus par le Cati (Centre Automatisé de Traitement de l'Information) Sicpa (Systèmes d'Informations et de Calcul pour le Phénotypage Animal), et des données provenant d'autres sources de stockage.

**Mots clés :** ontologies, diffusion, interopérabilité, labellisation

## Introduction

En optant pour une mutualisation de leurs outils de gestion de données d'élevage et d'expérimentation dans leurs Unités et Installations Expérimentales (UE/IE) et un stockage des données dans des bases communes par espèce ou par nature de données, les Départements Génétique Animale (GA) et Physiologie Animale et Systèmes d'Élevage (Phase), ont franchi une première étape pour rendre les données collectées interopérables. Mais aujourd'hui nombre de projets scientifiques visent à tirer parti d'une agrégation de données de sources hétérogènes, demandant de monter d'un cran leur niveau d'interopérabilité pour pouvoir mettre en relation des données phénotypiques comparables et traiter en masse de grandes quantités d'informations. Les ontologies constituent en cela un moyen de garantir que, pour un terme donné, tout le monde parle bien de la même notion, quelle que soit la provenance de la donnée. De même pour une variable donnée, on doit pouvoir préciser la façon dont elle a été mesurée et acquise.

## De l'interopérabilité des systèmes d'informations Sicpa à celle des jeux de données au-delà des UE/IE

Le Larousse nous donne cette définition de l'interopérabilité : « Compatibilité des équipements, des procédures ou des organisations permettant à plusieurs systèmes, forces armées ou organismes d'agir ensemble : interopérabilité des forces de l'Otan. Informatique : capacité de matériels, de logiciels ou de protocoles différents à fonctionner ensemble et à partager des informations : interopérabilité des réseaux téléphoniques. ». Cette notion transversale, qui permet à divers outils de pouvoir communiquer entre eux, suppose que l'on sache comment et pourquoi ils peuvent et doivent fonctionner ensemble. En ce qui concerne ses SI, le Cati Sicpa a choisi, pour ses nouveaux projets, de développer, quand c'était possible, des systèmes d'informations multi-espèces interopérables avec ses autres SI et notamment les SI espèces ; cela permet de mutualiser des développements et d'éviter les redondances d'informations dans les bases de données ainsi que des saisies multiples pour les utilisateurs. L'intérêt d'une telle démarche est donc de croiser de façon transparente des informations stockées dans des bases de données différentes et les informaticiens du Cati mettent en œuvre cette démarche au travers de webservices qui gèrent les échanges entre SI.

Mais les données stockées dans nos bases peuvent-elles être utilisées par d'autres chercheurs, dans d'autres contextes que celui de leur acquisition, dans d'autres organismes et d'autres pays ? En posant cette question, on ne présume en rien de leur diffusion mais on s'interroge sur leur conformité par rapport à des standards techniques, leur compatibilité avec d'autres applications, leur capacité à être agrégées avec des données de sources différentes, souvent hétérogènes.

1 UMR Physiologie, Environnement et Génétique pour l'Animal et les Systèmes d'Élevage, Inra, 35590 Saint-Gilles, France

2 UMR Génétique Physiologie et Systèmes d'Élevage, Inra, 31326 Castanet-Tolosan Cedex, France  
sicpa-contact@inra.fr

L'idée la plus simple consiste peut-être à définir une norme ou un ensemble de normes pour rendre possible l'exploitation conjointe des informations issues de différentes sources, alors qu'elles sont la plupart du temps générées avec des méthodes variées et qu'elles répondent souvent à des besoins spécifiques. Avoir un vocabulaire commun de définition des données, mieux préciser et mettre à disposition les métadonnées ou les méthodes et les conditions d'obtention, doivent permettre de faciliter le dialogue.

## Les ontologies – un moyen d'envisager l'interopérabilité des données

Comme précisé dans l'article de Meunier-Salaun et al. (2018), la piste des ontologies est aujourd'hui étudiée pour établir des liens entre des bases de données hétérogènes. Elles constituent un levier de valorisation intéressant en permettant un accès normalisé aux données. Pour préciser la définition du terme ontologie (« théorie de l'être ») on peut dire que pour un domaine spécifique, une ontologie est un référentiel univoque et partagé par l'ensemble des utilisateurs ; dans notre cas d'utilisation, les bases d'ontologie décrivent les caractères phénotypiques des animaux et leur environnement.


Mais comment établir des relations entre nos bases de données phénotypiques et les traits ontologiques décrits dans des ontologies d'animaux de rente comme Atol (ontologie des traits phénotypique des animaux d'élevage) et Eol (ontologie des paramètres environnementaux d'élevage) ? Aujourd'hui nous n'avons pas encore de réponse précise et plusieurs pistes sont à l'étude ; faut-il associer à chaque variable de nos SI un numéro de trait d'ontologie, doit-on envisager une affectation au moment des extractions de données ? La question reste ouverte mais la réponse passera certainement par une première étape d'analyse comparée qui conduira probablement à un enrichissement des ontologies car des études préliminaires montrent que certains caractères ne sont pas présents ou doivent être précisés plus finement. Dans d'autres cas, on a constaté qu'une variable de nos bases de données pouvait être associée à deux caractères d'ontologie correspondant à des notions différentes. Nous n'avons pas encore inventorié tous les cas de figure et nous ne savons donc pas encore comment ils pourront être traités. À titre exploratoire, un de nos SI propose désormais, à la demande de scientifiques, la saisie d'un numéro de trait d'ontologie lors de la définition d'une nouvelle variable et pour le nouveau projet sur les conditions d'ambiance, le lien avec les bases d'ontologie a été abordé dès les premières phases des enquêtes terrain réalisées préalablement au lancement du projet.

Pour avancer dans notre réflexion et notre analyse, plusieurs informaticiens du Cati collaborent avec des scientifiques spécialistes des ontologies dans le cadre de projets spécifiques : le Cati est notamment impliqué dans le projet Ahol (animal health ontology for livestock) via le SI Sanitaire et dans le projet D-Ont « Exploitation optimisée des bases de données phénotypiques - Des ontologies pour le partage d'information » dont le périmètre dépasse les bases de données Sicpa. Plus récemment, il a été sollicité dans le cadre du projet européen SmartCow, par Catherine Hurtaud, ingénieure de recherche à l'Unité Mixte de Recherche Pegase, dont un des objectifs consiste à étudier les possibilités d'adaptation de la liste des traits Atol à partir des caractères mesurés en routine sur les animaux des troupeaux expérimentaux bovins et d'adjoindre à certains des caractères décrits, leurs principales méthodes de mesures.

## Conclusion

Créer des passerelles entre nos bases de données élevages et les bases de données d'ontologie constitue un moyen de mieux valoriser les données phénotypiques de nos SI et de les engager dans un nouveau cycle de vie au-delà du suivi d'élevage ou de l'expérimentation qui les a fait naître. On rejoint en cela des enjeux clairs affichés dans le schéma directeur du numérique de l'Inra et dans le processus de labellisation des UE





autour de l'ouverture et du partage de données, d'outils et de pratiques. Cela ne veut pas pour autant dire que tout un chacun pourra venir interroger nos bases (Martignon et al., 2018) et toutes les données qui en font la richesse, mais que les moyens de l'interopérabilité au travers des ontologies seront testés et reportés dans nos SI. On peut aussi noter que, si les ontologies doivent être accessibles à l'international et donc les traits exprimés en anglais, il ne faut pas oublier que ce seront les techniciens animaliers qui collecteront ces données ; une traduction française s'impose pour qu'ils puissent communiquer sans ambiguïté avec les scientifiques.

Aujourd'hui les lignes directrices et les choix techniques ne sont pas encore bien établis ; pour notre collectif d'informaticiens, l'idée est d'avancer de façon coordonnée et réfléchie avec les différents acteurs concernés, car, en dehors des aspects techniques, les orientations et les choix stratégiques concernant la diffusion des données, ne relèvent bien évidemment pas de notre seul niveau de responsabilité. La réponse à apporter aux besoins d'interopérabilité des données sera un des défis de la nouvelle génération de notre Cati.

## Références bibliographiques

Journaux A, Heirman T, Reichstadt M, Coudert T, Robelin D, Chalier P (2018) Méthodes et outils informatiques du Cati Sicpa. *Le Cahier des Techniques de l'Inra*, N° Spécial phénotypage animal, pp. 27-36.

Martignon M, Allain D, Vignon X, Le Guerhier F (2018) Valorisation et diffusion des données. *Le Cahier des Techniques de l'Inra*, N° Spécial phénotypage animal, pp. 150-152.

Meunier-Salaün MC, Bugeon J, Fatet A, Hue I, Hurtaud C, Nédellec C, Vernet J, Reichstadt M, Le Bail PY (2018) Un outil au service de la standardisation des bases de données : les ontologies Atol/ Eol. *Le Cahier des Techniques de l'Inra*, N° 93.

Pasquet A (2015-2016) Vérification de la présence des caractères enregistrés dans le cadre d'expérimentations, ou dans les bases de données d'unités expérimentales (phénotypage, gestion zootechnique du troupeau) dans l'ontologie Atol. Mémoire de Master 1 Biologie, Agronomie, Santé à l'Umr Pegase, Inra.