



Organisation mondiale  
de la santé animale  
Fondée en tant qu'OIE

---

## **THEME TECHNIQUE II**

Progrès des Services vétérinaires grâce à la numérisation (gestion des données, systèmes d'information vétérinaires, Big Data, métalangage, intelligence artificielle, etc.)

**Prof Beatriz Martinez Lopez**

18/09/2024

**Thème technique II :**  
**Progrès des Services vétérinaires grâce à la numérisation (gestion des données, systèmes d'information vétérinaires, Big Data, métalangage, intelligence artificielle)**

**Prof Beatriz Martinez Lopez**

Directrice du Center for Animal Disease Modeling and Surveillance

## **Résumé**

L'énorme masse de données relatives à la santé animale produite journalièrement ainsi que les analyses de Big Data peuvent représenter une révolution quant à la manière dont nous traitons les questions de lutte contre les maladies et l'épidémiologie. Les avancées enregistrées en matière de génomique, de capteurs et de technologies de l'information permettent une caractérisation plus détaillée et plus précise de la santé animale. Dans le présent rapport, nous allons aborder la valeur que la numérisation peut apporter aux Services vétérinaires et encourager à progresser vers une épidémiologie vétérinaire de précision qui devienne un concept profitant des données à multiveaux sur la santé animale afin de mieux comprendre la dynamique de la maladie au sein d'une population et de concevoir des systèmes plus économiques portant sur la surveillance, la détection précoce et la lutte rapide contre les maladies des animaux d'élevage. Nous allons explorer les sources d'information, voir comment cela fonctionne et identifier les principaux défis à relever pour transposer ce concept en pratique.

Nous pensons que la numérisation peut non seulement contribuer de façon importante à l'amélioration du rendement en matière d'animaux d'élevage et de la durabilité du fait de la réduction de l'empreinte environnementale mais également donner un coup d'accélérateur aux avancées verticales et horizontales portant sur la façon dont nous prévenons et gérons les maladies des animaux d'élevage aux niveaux local et mondial. Toutefois, pour concrétiser cela, il faut avoir des avancées critiques et des changements quant à la façon dont nous collectons, normalisons, intégrons, partageons et utilisons les données. Il y a aussi la nécessité de construire des équipes interdisciplinaires de scientifiques, d'ingénieurs en informatique et /ou d'experts en matière de données qui travaillent en collaboration avec des vétérinaires et des experts d'autres domaines et qui sont centrés sur les questions de santé animale. Il est certain que cela comporte des défis mais il s'agit aussi d'une occasion extraordinaire pour les Services vétérinaires d'appliquer ces outils existant du numérique pour améliorer notablement la santé animale.

## Introduction

La collecte, la manipulation et l'analyse des données sont essentielles pour documenter des interventions intervenant en temps voulu et concevoir des programmes de surveillance et de contrôle aussi économiques que possible. La collecte de données relatives à la santé animale a été traditionnellement cantonnée à rassembler et surveiller certains indicateurs spécifiques censés communiquer des informations afin d'évaluer certains problèmes spécifiques. Toutefois, ces dernières décennies, la génération, la collecte et la possibilité de numérisation des données, de leur stockage et de leur analyse ont démarré en flèche pour élargir le champ des possibilités. De cette façon, de nos jours, de larges quantités de données multisectorielles sont produites quotidiennement et sont collectées par des producteurs, (tels que documents portant sur les échanges commerciaux, capteurs portables par les animaux), vétérinaires et laboratoires de diagnostic (documents sanitaires pour les animaux) ou systèmes de gestion et de suivi environnemental (température, humidité relative, vent, etc.), entre autres; elles peuvent être utilisées pour améliorer la gestion de la santé animale. Ces grands nombres de données constituent ce que l'on appelle le « Big Data », un énorme groupe de données extrêmement vaste, complexe, et divers qui ne peut pas être traité par des moyens traditionnels<sup>1,2</sup>.

Il est évident que le recours à ce Big Data peut révolutionner la façon dont nous appréhendons la gestion de la santé animale tout comme la collecte des paramètres relatifs à la reproduction dans l'élevage ou les documents relatifs à la mortalité ont aidé dans le passé à accroître la productivité et le rendement des animaux d'élevage. Nous sommes capables de comprendre les problèmes zoonosaires de façon plus précise et transverse et passer à des approches qui soient davantage proactives et adaptées. Toutefois, la gestion du (Big) data et l'application des analyses de Big Data ne sont pas simples et doivent se confronter à plusieurs obstacles, traditionnellement définis par ce qu'on appelle les trois V: 1) volume – reflétant l'accroissement en taille de ces jeux de données, 2) vitesse – le rythme rapide et souvent exprimé en temps réel auquel les données sont actualisées et 3) variété – c'est à dire les différents types de données collectées au sein des diverses sources et échelles spatiotemporelles<sup>3</sup>. De plus, il est possible d'ajouter deux autres V importants : la véracité et la valeur, afin de tenir compte de la nécessité de disposer de données intégrées et fiables et d'avoir des résultats exploitables issus de l'analyse<sup>4,5</sup>. En outre, d'autres V, allant jusqu'à sept<sup>6</sup>, dix<sup>7</sup> ou même 42<sup>8</sup> ont également été suggérés afin de décrire plus en détail les caractéristiques et les propriétés du Big Data. Toutefois, il est plus facile désormais de surmonter ces difficultés avec la disponibilité des ressources informatiques qui facilitent la capture des données, le stockage et le traitement de grands volumes de données, l'informatique en nuage " cloud computing", les avancées en matière d'automatisation et l'apprentissage machine ainsi que l'élaboration d'outils analytiques et de plateformes intégrées. De ce fait, l'intérêt porté à l'analyse de Big data s'est accru de façon exponentielle afin de mieux gérer la santé animale. Toutefois, les Services vétérinaires en de nombreuses parties du monde ne disposent pas de la capacité à collecter les données de façon numérique et standardisée et ils peuvent ne pas avoir les infrastructures numériques nécessaires ainsi que le personnel spécialisé pour collecter les données de façon numérique et standardisées et peuvent aussi ne pas disposer des infrastructures informatiques et du personnel spécialisé afin de pouvoir correctement manipuler ces données. C'est un obstacle réellement important, mais il s'agit de quelque chose qui peut être relativement facilement appréhendé. De fait, la création d'une infrastructure visant à collecter et à manipuler ces données (smart phones, tablettes, ordinateurs portables, serveurs, etc.) tout en regroupant une poignée d'ingénieurs, de programmeurs et de spécialistes informatiques ou de données visant à aider les Services vétérinaires ne serait pas une opération très coûteuse et pourrait se révéler être un investissement incroyablement rentable. Il y a un manque de vétérinaires au niveau mondial mais avec l'aide de ces autres professionnels nous pourrions mieux utiliser le « temps des vétérinaires » puisque ces autres personnes pourraient faciliter et accélérer la collecte, l'accès et la visualisation des données pertinentes afin que les vétérinaires puissent se concentrer pleinement sur l'évaluation de ces informations et prendre les meilleures décisions en matière de santé animale.

Le présent rapport vise à donner un aperçu sur les aspects suivants :

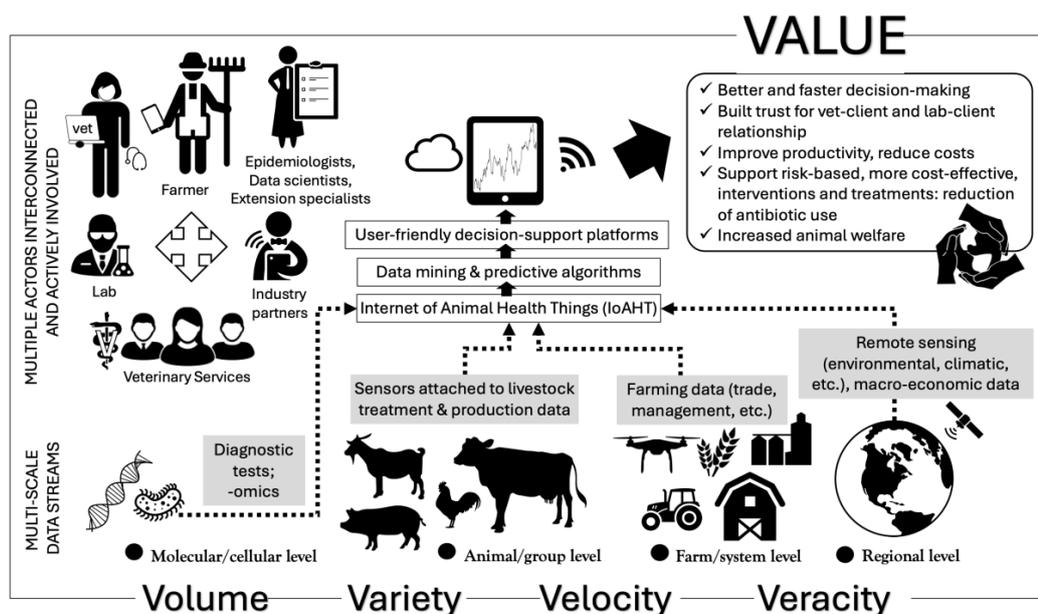
- 1) **Évolution et statut actuel du numérique en matière de santé animale**
- 2) **Considérations relatives aux données** : types de données, sources de données, normes relatives aux données, qualité des données, etc. nécessaires pour mieux prévenir et contrôler les maladies des animaux producteurs d'aliments
- 3) ce que sont, selon nous, les cinq « **Défis principaux pour le numérique en matière de santé** » afin d'accélérer la révolution numérique en matière de santé animale.

## 1. Statut actuel du numérique en matière de santé animale

Les applications et les défis du numérique dans le domaine de la santé animale et liés aux analyses du Big Data ont récemment été évoqués dans plusieurs études et articles d'opinion. Certains d'entre eux portent sur les opportunités et les défis de l'Internet des objets en matière de santé animale (IoAHT)<sup>9</sup>, la valeur de l'utilisation du séquençage de nouvelle génération (NGS), permettant le séquençage ADN/ARN et la détection de mutation en très peu de temps avec l'association du Big Data pour la médecine de précision chez les chevaux et autres espèces animales<sup>10</sup>, la valeur que cela représente pour le bien-être animal<sup>11,12</sup> ou pour améliorer la prise de décision en matière de santé animale<sup>13,5</sup>. Toutefois, le Big Data en matière de santé animale et d'épidémiologie vétérinaire a été énormément infra-utilisé en pratique<sup>14</sup>. Les données continuent à être présentées et utilisées de façon séparée, la plupart du temps avec une mauvaise communication, voire une absence de communication entre les différents segments impliqués et les défis sont importants à chaque étape de la chaîne, de la collecte des données jusqu'à leur analyse et à la production de résultats complets. Résoudre tous ces problèmes serait extrêmement profitable pour la prévention et le contrôle des maladies.

Un certain nombre de disciplines de « précision » est apparu pour profiter des possibilités offertes par la numérisation et les analyses du Big Data. C'est le cas du Precision livestock farming (PLF) (agriculture et élevage de précision), impliquant l'interconnexion d'une large gamme de technologies en temps réel afin de surveiller les paramètres cibles ou clefs et ainsi, identifier et notifier tout évènement marquant intéressant pour les élevages de façon à optimiser les différents aspects de la production d'animaux d'élevage et ceci a pu se faire avec une réussite relative<sup>10,15</sup>.

De la même façon, nous pouvons exporter les mêmes principes pour la santé animale et le contrôle des maladies en profitant d'outils tels l'apprentissage machine et autres méthodologies issues de l'intelligence artificielle (IA). Cette approche pourrait contribuer à une détection précoce des épisodes de maladies et améliorer la gestion de la santé animale en comprenant mieux la dynamique de l'infection ainsi qu'une identification et une détection précoces des schémas et des facteurs liés à la maladie en combinant les méthodes traditionnelles avec les analyses du Big Data. Ceci pourrait se révéler fort utile pour aider les Services vétérinaires à mieux concevoir et mettre en œuvre les programmes de prévention et de contrôle. En fait, le concept d'épidémiologie de précision ou de « santé publique de précision » a déjà été exploré en médecine humaine<sup>16</sup> définissant cette dernière comme « *proposant la bonne intervention à la bonne population au bon moment* »<sup>17</sup>. En santé animale, l'épidémiologie vétérinaire de précision pourrait être obtenue en intégrant les données multisectorielles très variées déjà générées chez les animaux d'élevage (à savoir, diagnostics des animaux, gestion, biosécurité, indices de production, échanges commerciaux, économie, etc.). A l'aide de ces données à haute résolution, nous pouvons mieux connaître le problème épidémiologique et aider à une sélection plus rapide d'interventions adaptées à des groupes spécifiques d'animaux, d'élevages ou de systèmes de production, L'épidémiologie vétérinaire de précision ne vient pas remplacer l'épidémiologie « classique » mais profite de la façon dont les technologies numériques et les nouveaux outils analytiques dont on dispose permettent des évaluations davantage proactives, rapides et adaptées de la santé des populations, des risques de maladies et des vulnérabilités présentées par l'élevage. Nous pensons que cette approche va représenter une grande valeur pour les producteurs d'animaux d'élevage, des vétérinaires privés et des Services vétérinaires permettant une information plus précise et plus rapide et constituer une aide à la décision (Figure 1).



**Figure 1. Schéma illustrant la façon dont le numérique peut faciliter l'intégration de flux de données multiples et comment la communication au sein des parties prenantes peut générer de la valeur en matière de santé animale.**

## 2. Considérations des données

La grande variété des données portant sur la santé animale est l'un des points forts (mais également le principal défi) pour permettre d'arriver à une numérisation totale pour les Services vétérinaires. La collecte des informations portant sur les animaux et les élevages (à savoir, numérisation de l'agriculture) s'est notablement spécialisée et diversifiée au fil des ans, notamment au sein des systèmes de production à grande échelle des pays à hauts revenus. Il est facile aujourd'hui d'obtenir des informations au sein des troupeaux, des animaux ainsi qu'au niveau moléculaire mais également à plus large échelle, au-delà même des troupeaux et des collectifs, à partir d'une immense variété de ressources qui collectent, stockent et partagent les données multiples : dispositifs automatiques de données sur/au sein des animaux, les capteurs dans les élevages, les enregistrements provenant de différents programmes de surveillance mis en place pour le troupeau, les laboratoires, les institutions, différentes entreprises au sein de la filière de l'élevage, les rapports obligatoires ou les programmes, etc. Ceci augure d'une forte capacité à caractériser et à comprendre un système (les risques, points faibles et points forts qu'il comporte) pour permettre ainsi une grande adaptation des interventions. Toutefois, la richesse de ces données, de ces sources de données et de ces types de données représentent également la principale difficulté pour pouvoir mettre en œuvre de façon opérationnelle l'épidémiologie vétérinaire de précision. C'est une tâche complexe d'intégrer, de configurer, de visualiser et de fournir un accès en temps voulu aux parties prenantes à toutes ces données, étant donné que ces données sont exclusives et sont collectées, stockées et manipulées par diverses entités (à savoir, gestion de production, laboratoires de diagnostic, cliniques vétérinaires, entreprises pharmaceutiques, etc.) et sans suivre une structure de données commune, des définitions de cas ou des normes relatives aux données. De plus, dans les pays ou dans les zones à revenu faible ou intermédiaire où se trouvent de petits exploitants, la plupart de ces données ne sont toujours pas collectées ou, si elles le sont, elles le sont sur papier (enquêtes épidémiologiques et autres questionnaires, par exemple) ce qui rend plus compliqué l'intégration, le traitement et l'utilisation et va, sans nul doute, retarder l'accès aux données et poser un problème pour les applications en temps réel. L'utilisation de tablettes ou de smart phones pour que les Services vétérinaires puissent collecter des informations sur le terrain peut réduire les erreurs pour entrer les données et accélérer l'utilisation de ces données pour aider à une prise de décision plus rapide. Toutefois, cela ne sera pas suffisant. La recherche et les politiques en la matière doivent avoir pour objectif de progresser vers un processus de numérisation plus équitable et inclusif en facilitant l'accès aux possibilités du numérique pour les petits exploitants et des Services vétérinaires dans des milieux à faible revenu. Nous aborderons ci-après les principales sources de données et les types de données relatives à la santé animale aux niveaux moléculaire, de l'animal, de l'exploitation et de la région.

## 2.1. Information aux niveaux moléculaire/cellulaire

Les informations moléculaires représentent un important pilier du diagnostic, du suivi et de la lutte contre les maladies infectieuses. Le diagnostic et le séquençage PCR sont réalisés régulièrement au sein de plusieurs segments d'animaux d'élevage afin de détecter et d'identifier des souches spécifiques, émergentes circulant au sein d'un élevage ou afin de repérer de nouvelles infections. Avoir recours à des informations moléculaires est une pratique bien établie en cas d'enquête épidémiologique afin de comprendre la source d'un foyer et sa répartition au niveau de la population, à l'aide, par exemple, de la phylogéographie<sup>18</sup>.

De nouvelles technologies moléculaires, telles que le séquençage de nouvelle génération (NGS) et l'analyse de l'expression des gènes (un processus par lequel l'information encodée dans un gène est transformée pour en faire un produit, une protéine ou une fonction finale par un processus de transcription et de traduction) sont de plus en plus présentes en raison d'une réduction des coûts et d'une plus grande accessibilité ce qui facilite la production et la mise à disposition d'informations génomiques massives en matière d'épidémiologie. Ceci a abouti à l'émergence de différentes disciplines regroupées sous le terme d'omiques (protéogénomiques, métabolomiques, métagénomiques, transcriptomiques, etc.), des méthodologies à haut débit pouvant mesurer toutes les mêmes molécules simultanément pour donner une vue holistique d'un processus moléculaire. L'application de ces technologies à l'épidémiologie suppose une révolution quant à la façon dont les agents pathogènes et les hôtes peuvent être caractérisés. Nous pouvons, par exemple, identifier des traits spécifiques liés à la résistance aux agents pathogènes<sup>19-21</sup>, aux récepteurs cellulaires clefs chez des individus spécifiques<sup>22,23</sup> ainsi que les voies moléculaires de l'infection<sup>24-26</sup>; la forte persistance des infections, la virulence, les interactions hôte-agent pathogène ainsi que la pathogénicité, la colonisation et les mécanismes de survie<sup>27,28</sup>; comprendre la diversité génétique des agents pathogènes ainsi que leur interaction avec les communautés de flores microbienne; évaluer la résistance aux antimicrobiens et aux gènes résistants qui y sont associés<sup>29-30</sup>. Les omiques aident également à identifier les biomarqueurs des infections ou de la maladie<sup>31,32</sup>, qui peuvent servir à surveiller des collectifs et à détecter la maladie de façon précoce alors qu'elle sévit sans être remarquée<sup>33</sup>.

Ces nouvelles technologies moléculaires donnent une image plus précise des relations existant entre l'hôte et l'agent pathogène et peuvent constituer un changement de paradigme portant sur la façon dont nous étudions les problèmes biologiques. Chez l'être humain, nous avons vu que la médecine, met déjà en œuvre ces outils dans la médecine de précision afin de concevoir des traitements personnalisés pour les individus en tenant compte des caractéristiques spécifiques des hôtes et des agents pathogènes<sup>16</sup>. Le même principe peut être extrapolé à l'épidémiologie vétérinaire afin de personnaliser les programmes de contrôle et d'éradication ou pour rechercher l'origine et l'extension d'un foyer. Les foyers dus au même agent pathogène et à la même souche peuvent présenter des dynamiques différentes au sein de troupeaux divers ce qui explique l'importance des mesures de lutte contre la maladie qui prend en compte les interactions spécifiques existant entre l'hôte, l'agent pathogène et l'environnement dans un élevage donné. De ce fait, les informations que peut offrir le séquençage nouvelle génération (NGS) ont un énorme potentiel en matière de surveillance épidémiologique, détection des foyers, et de lutte contre l'infection. Toutefois, la principale limitation quant à sa mise en œuvre par les Services vétérinaires vient de son coût élevé et d'une gestion complexe de l'information. Il reste à faire en sorte que ce séquençage soit d'un prix plus abordable et à élaborer des systèmes en mesure de gérer et d'analyser de façon plus efficiente l'énorme volume de données qu'ils produisent.

## 2.2. Information au niveau de l'animal et de l'élevage

Les capteurs sont utilisés de nos jours pour collecter des données soit issus de l'environnement dans lequel les animaux sont placés, soit provenant de l'animal ou de ses excréments (lait, salive, par exemple...). Le développement des technologies par capteurs a énormément progressé ces dernières années et nous avons désormais le potentiel pour mesurer pratiquement tout paramètre de l'animal ou de l'élevage pouvant être nécessaire pour une enquête épidémiologique, tel que la fièvre, le mouvement de l'animal, la nourriture ingérée et l'eau consommée<sup>10,34,14,35</sup>.

Ces technologies se sont infiltrées à des rythmes différents dans la santé animale en fonction de la filière considérée. L'industrie laitière a été une de celles qui s'est très tôt prononcée pour la génération, la collecte et le recours au Big Data par le biais de dispositifs et de systèmes cyber-physiques, tels que des robots ou des systèmes de traite automatisés afin de mieux détecter et gérer les problèmes liés à la reproduction, identifier la mammite, surveiller la qualité du lait ou la boiterie<sup>6</sup> et détecter les substances produites par les animaux: progestérone, composés organiques volatils (corps cétoniques, éthanol, méthanol, etc.)<sup>14,41</sup>. La filière avicole tout comme la filière porcine ont fait d'importants progrès dans la mise en œuvre des technologies des capteurs, bien que leur adoption demeure assez faible, surtout en raison des coûts<sup>11,42</sup>. Nous sommes capables désormais de surveiller l'environnement (température, humidité et dioxyde de carbone et

qualité de la ventilation)<sup>43</sup>; ou des marqueurs de maladies en temps réel. Pour ne citer que quelques exemples, il est possible d'avoir recours à l'imagerie thermique à infrarouge pour surveiller les contraintes thermique<sup>44</sup>, les accéléromètres pour l'activité<sup>45</sup>, les marques auriculaires pour le comportement<sup>46</sup>. La température corporelle moyenne ainsi que les relevés de mobilité ont servi à la détection précoce de l'influenza aviaire chez les poulets<sup>47</sup>. Des microphones ont également été utilisés pour surveiller la toux et détecter les problèmes respiratoires chez les porcs<sup>48</sup>; des cellules photoélectriques ont servi à détecter la boiterie<sup>49</sup>, des systèmes de caméras ont été utilisés pour évaluer le poids du corps et les changements en matière de comportement et d'activité<sup>50,51</sup>.

Bien que l'utilisation de capteurs ait connu une croissance exponentielle au sein du PLF<sup>42,52,53</sup>, avoir recours à ces données à des fins zosanitaires reste encore à développer totalement. En fait, la plupart de ces technologies de détection doivent se voir plus largement adoptées dans la filière de l'élevage, étant donné qu'actuellement elles ont été utilisées comme preuve de concept ou comme des applications isolées et indépendante. En termes d'épidémiologie vétérinaire, l'intégration et l'usage en synergie de données issues de capteurs combinés à d'autres données (à savoir relevés de santé, informations relatives à la performance de production, à la biosécurité, aux vaccinations, aux traitements, aux échanges commerciaux et autres pratiques managériales dans l'élevage), apporterait une valeur ajoutée et contribuerait à faire avancer la gestion de la santé animale.

### 2.3. Informations au niveau régional

Les informations à ce niveau comportent une couche hétérogène composée de données venant de sources très différentes. En premier lieu, nous générons des recensements ou des données relatives à la population simplement en collectant les informations décrites ci-dessus pour chaque animal/élevage et qui sont ensuite rassemblées pour caractériser la population dans son ensemble (à savoir, système de production association d'élevages, région, etc.). Les données portant sur le recensement aident à identifier les tendances, à évaluer le risque de contact avec des agents pathogènes et à prendre des mesures de biosécurité en étant informés et à avoir une meilleure planification en cas de riposte d'urgence au niveau régional. Toutefois, il existe plein d'autres ressources pouvant également servir à caractériser les collectifs. Par exemple, l'utilisation régulière des diagnostics aboutit à ce que les laboratoires et les cliniques vétérinaires recueillent un nombre important et un volume croissant de données de diagnostic et relatifs aux maladies pouvant être utilisés pour évaluer les tendances de la population. Toutefois, les données au niveau de l'animal et de l'élevage relèvent de la propriété privée et ne sont pas souvent accessibles pour les gens en-dehors de ceux qui opèrent au niveau de l'élevage. Partager des données est une chose rare et compliquée en raison des questions de confidentialité et du fait d'un manque de systèmes correctement conçus pour collecter, intégrer et renvoyer les informations de manière compréhensible. Toutefois, le partage de documents sanitaires groupés ou désidentifiés s'est révélé extrêmement précieux pour informer les producteurs des tendances générales et peut se révéler utile pour améliorer la sensibilisation lorsque surviennent des changements dans les schémas et l'épidémiologie de certaines maladies, notamment pour les maladies endémiques<sup>54</sup>. Nous pouvons aussi retrouver des traces de foyers de maladies soumises à déclaration dans les bases de données publiques nationales ou internationales (Maladies animales soumises à déclaration WAHIS, <https://wahis.woah.org>).

D'autres types d'informations précieuses sont constituées par des relevés des échanges commerciaux d'animaux ainsi que sur les mouvements des camions, les statistiques d'abattage, les facteurs environnementaux et climatiques, etc. Les applications peuvent en être multiples. L'analyse de réseau portant sur les mouvements des animaux peut aider à mesurer les risques par le biais des échanges commerciaux d'animaux<sup>55,56</sup>; un regroupement spatial et temporel des cas qui est utile pour identifier les zones et le moment présentant les plus grands risques d'apparition de la maladie<sup>57,58</sup>; les modèles de transmission de la maladie peuvent contribuer à simuler la propagation de la maladie et à identifier les interventions<sup>59</sup> les plus efficaces par rapport à leur coût. Les données environnementales, géographiques et climatologiques (températures, précipitations, indices relatifs à la végétation, marqueurs environnementaux, etc.) sont des modulateurs connus de risques de maladie largement utilisés dans les études de recherche exploratoires, la modélisation et la cartographie des maladies. Actuellement, la collecte des données climatologiques est réalisée par le biais de stations météorologiques, par des stations d'observation au sol et peut être limitée dans certaines zones. De plus, l'imagerie visible et infrarouge par satellite a amélioré au fil du temps la résolution spatio-temporelle et donné différentes mesures telles que le coefficient de réflexion du rayonnement qui peut être transformé et servir de marqueurs environnementaux et climatiques pour modéliser le risque d'urgence face à la maladie<sup>60</sup>. La haute résolution et la fréquence pratiquement en temps réel de la production de données météorologiques permettent de les utiliser à des fins de surveillance, d'évaluation rapide des zones à risques ou pour prévoir la répartition de la maladie.

Les nouvelles technologies de communication offrent également des ressources nouvelles de collecte d'information. Les données Internet peuvent être utilisées pour alerter sur l'émergence de maladies avant que les autorités sanitaires en aient connaissance<sup>61,62</sup>. En médecine humaine, différentes tentatives ont été faites afin d'utiliser les interactions entre les personnes sur l'Internet afin de détecter l'émergence de l'épidémie. Par exemple, Google Flu Trends a peut-être été l'une de ces toutes premières applications<sup>63</sup>, et au cours de ces dernières années, le suivi des réseaux sociaux a servi d'aide à la surveillance, à la détection d'évènement, à la pharmacovigilance, aux prévisions, au suivi des maladies et à l'identification géographique des maladies<sup>64,65</sup>. Différentes stratégies ont été employées pour surveiller les tweets, par exemple et les utiliser pour détecter l'émergence d'une maladie plutôt que par le biais de rapports officiels.

Les données de ce niveau offrent d'immenses possibilités en raison de la variété des données mais l'accès et l'intégration à toutes ces sources ne sont pas toujours faciles. Les informations portant sur l'élevage sont généralement limitées aux systèmes de production intégrés, aux associations d'éleveurs ou aux plans d'amélioration collective de la qualité (contrôle du lait chez les vaches laitières par exemple); cela est fait dans des buts bien précis, à différents degrés en matière de production animale, rarement en temps réel et il y a peu de partage. Les données au niveau de la population collectées en dehors de l'élevage présentent des problèmes d'intégration et un manque de normalisation. Certaines sources peuvent offrir de grands volumes de données avec une bonne périodicité, d'autres peuvent être plutôt incomplètes et rares. Cette hétérogénéité complique la cartographie des données, la manipulation et la modélisation des données de façon appropriée afin d'extraire la valeur.

### **3. Les défis posés par la numérisation pour la santé animale**

Les progrès accomplis par la numérisation pour aller vers une large utilisation et une mise en œuvre par les Services vétérinaires, font qu'il faut s'attaquer à ce que nous pensons être les cinq principaux défis posés par la numérisation en matière de santé animale

#### **3.1. Faciliter l'accès et l'intégration de données à plusieurs niveaux**

L'émergence, la propagation et l'expression de la maladie sont modulés par de multiples facteurs de nature très diverse. Si nous aspirons à une meilleure compréhension et donc à une meilleure prédictibilité des maladies, une approche holistique intégrant toutes les informations potentiellement en liaison avec l'épidémiologie des maladies et rassemblant tous les acteurs impliqués pourrait mieux refléter la chaîne de causalité et leurs interactions. Toutefois, en pratique, collecter et combiner les sources de données à plusieurs niveaux n'est pas chose fréquente. Même au sein d'un même niveau, les données sont conservées de façon dispersée, collectées généralement par différentes entreprises ou services et sont présentées dans des formats qui sont généralement difficiles à intégrer. En outre, les entreprises peuvent s'approprier l'enregistrement des données et ne pas toujours faciliter l'utilisation des données en dehors de leur système. Ce manque d'interopérabilité est l'un des problèmes le plus souvent évoqué à propos de la gestion du Big Data<sup>66</sup> pour les animaux et ceci porte atteinte non seulement à l'utilisation des informations, mais également pour explorer et en déduire de nouvelles applications. A la place, les résultats potentiels générés par l'interconnexion des données devraient se comprendre comme une valeur additionnelle peut-être plus importante et cette idée doit être transmise à l'utilisateur final. Des services et systèmes nouveaux peuvent découler de ce besoin et servir de plateformes véhiculant divers flux d'informations. Différentes initiatives ont démarré en agriculture recherchant la standardisation et l'interopérabilité entre les formats des données pouvant servir d'exemples (OASIS, 2016; OGC, 2016; GODAN, 2013), mais peu de travail a été réalisé au sein des animaux d'élevage jusqu'à ce jour<sup>66</sup>. Même en combinant des données similaires (informations en termes de diagnostic), le problème est parfois le manque de définitions de cas communs ou les disparités existant en matière de protocoles de collecte. En dépit des tentatives qui ont été menées à différents niveaux : signes cliniques<sup>67</sup>, données de diagnostic<sup>68</sup>; les normes de médecine vétérinaires sont très loin derrière celles de la médecine humaine. Il y a donc un besoin urgent de créer des normes portant sur la collecte de données relatives à la santé animale ou, du moins, de fournir des définitions de cas qui soient claires, des dictionnaires de métadonnées et de données afin de pouvoir être réellement capables de combiner, d'intégrer et de comparer toutes ces multiples sources d'information et de pouvoir en extraire davantage de valeur.

Intégrer des informations provenant de différents élevages soulève aussi des questions liées à la confidentialité et à la sûreté. L'exploitation des données ne doit pas être perçue comme pouvant potentiellement nuire à l'utilisateur, sinon, les éleveurs pourraient être réticents à partager leurs informations et c'est tout le système qui verrait son potentiel limité. Ceci peut être moins important pour les maladies endémiques quand tous les éleveurs souhaitent lutter contre une maladie ou l'éradiquer. En fait, les études comparatives et l'élaboration de « listes blanches » indemnes de maladies peuvent constituer un stimulus pour rejoindre un programme. Toutefois, en cas d'applications de pénalités ou de restrictions aux échanges commerciaux, la

confidentialité peut être une préoccupation limitant la capacité à détecter de nouveaux foyers ou des maladies émergentes. Il y a différentes méthodes d'apprentissage machine qui peuvent contribuer à surmonter ce problème. Papst et al.<sup>69</sup> ont proposé un aperçu des analyses de données garantissant la confidentialité. Parmi ces solutions, l'apprentissage fédéré est un type d'apprentissage machine qui a suscité un intérêt comme étant une solution prometteuse puisqu'il permet de partager des données distribuées et un apprentissage entre des machines sans avoir à transférer les données brutes. Au lieu de partager les données d'un élevage dans un système collectif, l'apprentissage fédéré peut collecter et traiter les données localement et partager le modèle en gardant les données sécurisées. Cette technologie s'est infiltrée dans divers domaines de ce qu'on appelle « l'agriculture intelligente » mais son utilisation pour les applications vétérinaires n'en est qu'à ses prémices.

Le problème de l'intégration a une importance toute particulière lorsque nous souhaitons introduire des données relatives aux omiques. Nous savons que les interconnexions entre les différentes disciplines « omiques » avec des données provenant d'autres niveaux (flux de données, par exemple) ont une grande pertinence pour comprendre la dynamique et les risques de la maladie. Toutefois, la gestion et l'intégration des omiques est particulièrement problématique en raison du grand volume et de la dimensionalité des données ainsi que du fait de leur hétérogénéité même au sein des omes, en raison de différentes échelles ou types de données (quantitatives, qualitatives, etc.). Différentes approches ont été proposées pour une analyse multi-omique: analyse statistique, analyse de réseau, analyse supervisée, etc.<sup>70-72</sup>. Elles sont aussi limitées fréquemment par la faible taille de l'échantillon en raison du coût des procédures des omiques, étant donné que la plupart des approches d'intégration nécessitent un grand volume de données. Des initiatives ont été entreprises afin de stocker et de standardiser les données métagénomiques pour les animaux<sup>73</sup> et il existe plusieurs bases de données publiques contenant des voies de référence pouvant faciliter la reconstruction et la prédiction des voies métaboliques<sup>74</sup>. Toutefois, il y a encore d'autres étapes qu'il faut franchir pour aller vers une intégration plus fonctionnelle. Il nous faut donc des **solutions à niveaux multiples conduites par les données et des cadres de décision** pouvant s'attaquer à certains de ces problèmes à échelles multiples et touchant la confidentialité

### 3.2. Comblent l'écart entre la disponibilité des données et leur utilisation réelle

Bien que nous ayons déjà collecté beaucoup de données et ayons la capacité d'en collecter encore davantage, leur utilisation se réduit souvent à de simples statistiques descriptives ou se limite à des aspects spécifiques de la production animale et des diagnostics des agents pathogènes<sup>75,76</sup>. Ceci veut dire que de nombreuses informations collectées ne sont pas utilisées et que nous n'extrayons pas la totalité du potentiel que présentent ces données qui ont coûté si cher à collecter. Les accéléromètres, par exemple, sont conçus et surtout utilisés pour détecter l'œstrus chez les vaches laitières mais cela peut aussi contribuer à identifier les animaux malades de façon précoce car ces derniers sont moins enclins à se déplacer. Il est donc important de souligner l'importance d'une **collecte des données à finalités multiples**. Une approche holistique de la collecte de données et une large variété d'indicateurs permettront de proposer des modèles plus précis et augmentera notre capacité à comprendre et à prédire des maladies.

Toutefois, il n'est pas suffisant de collecter davantage de données. Quantité ne veut pas dire qualité et disposer de beaucoup de données ne signifie pas forcément que ces données sont représentatives et fiables. De fait, des progrès doivent également être accomplis pour filtrer des données qui sont plus valables, informatives, représentatives, etc. ainsi que celles qui sont le plus facile à collecter et économiquement viables en fonction des objectifs. Le premier souci porte sur la qualité du processus de collecte lui-même. De nombreux liens peuvent intervenir dans la chaîne du processus de collecte des données et, si le personnel ou l'infrastructure qui y sont impliqués ne sont pas prêts à gérer les données, des pertes et des distorsions d'information peuvent survenir. Même lorsque les bonnes informations sont collectées, les données peuvent présenter des lacunes en fonction de la façon dont elles ont été collectées. Lorsque des informations manquantes sont dues à un facteur connu (congelés, jour de la semaine, résolution insuffisante par exemple, etc.), ces manques peuvent être quelque peu atténués par la prédiction, l'imputation et les maqueurs<sup>72,77</sup>. Toutefois, lorsque l'information manque en raison d'un schéma non aléatoire ou d'omissions systématiques, l'imputation n'est pas appropriée et la situation plus dure à compenser. L'apprentissage machine est particulièrement sensible à cet aspect, étant donné qu'il repose sur de la formation ou des données existantes prises comme une vérité de terrain et que cela peut fausser la représentation du système et donner de mauvaises estimations en restreignant la capacité des modèles d'IA à extraire l'information.

Afin de transformer les données en un élément informatif, une expertise spécifique, de la technologie et de l'analyse sont nécessaires. Des approches interdisciplinaires avec des professionnels possédant les aptitudes appropriées et l'expertise dans le domaine considéré à chaque maillon de la chaîne constituent la meilleure façon de procéder mais cela nécessite également la nécessité d'établir de bonnes voies de communication, de partage d'information et de retour d'information. Le projet « Pig Data » a fourni un exemple

illustrant la mise en œuvre de Big Data par le biais d'une approche multidisciplinaire au sein de la production des animaux d'élevage<sup>78</sup>. Des initiatives, telles que Digital Innovation Hubs, dans lesquelles les plateformes (une université par exemple) ont pour objectif d'aider et d'assister les parties à proposer des applications numériques peuvent aider à créer des groupes interdisciplinaires, comme cela se fait en élevage de précision<sup>79</sup>. Il est possible d'appliquer une approche similaire pour améliorer la numérisation dans le secteur de la santé animale et faciliter l'intégration de l'épidémiologie vétérinaire de précision dans les Services vétérinaires.

### **3.3. Elaborer de nouveaux algorithmes spécifiquement adaptés à la santé animale**

Généralement meilleures sont la quantité et la qualité des données introduites dans un modèle et meilleurs sont les résultats de ce modèle (plus grande précision, par exemple). Les stratégies de lutte par avertissement précoce sont donc facilitées par l'incorporation de données à niveaux multiples, mais les structures à niveaux multiples représentent également des défis analytiques en raison de structures complexes et des dimensionalités. Les algorithmes de l'apprentissage machine sont sans nul doute des outils puissants et ils constituent la meilleure approche pour exploiter ces entrées<sup>80</sup>, puisqu'ils présentent des avantages pour gérer des ensembles de données complexes, importants et extraire les valeurs de façon plus semi-automatisée. Lorsqu'ils sont appliqués à un flux constant de données issu de l'élevage, ces algorithmes peuvent être utilisés pour surveiller le système et détecter les anomalies pouvant indiquer une urgence due à une maladie avant que celle-ci puisse être remarquée par les éleveurs. L'adoption d'algorithmes d'apprentissage machine a connu une hausse exponentielle ces dernières années dans différentes disciplines dont la santé mais leur application en matière de santé animale reste limitée<sup>81</sup>. Guitian et al.<sup>82</sup> ont représenté de façon résumée les algorithmes d'apprentissage machine appliqués à des fins de santé animale se répartissant principalement en quatre domaines : diagnostic, évaluation du risque de mortalité et de morbidité, prédiction et surveillance des foyers de maladie ainsi que la politique et la planification sanitaire. Ils incluent de très larges algorithmes et applications : réseaux neuraux pour la reconnaissance par image, fouilles de textes afin d'extraire les valeurs des registres, arbres de classification afin d'étudier la prise de décision, priorisation des échantillons, détection précoce des maladies, détecter les réservoirs potentiels, etc. L'algorithme optimal d'apprentissage machine dépend souvent énormément de l'objectif et de la composition de l'ensemble de données. Une approche idéale d'épidémiologie vétérinaire de précision doit intégrer différents algorithmes et des analyses multiples afin de mieux aider à la prise de décisions en matière de santé animale. Le problème est que les techniques existantes sont surtout élaborées pour des données à un seul niveau, il faut donc adapter les algorithmes d'apprentissage machine à la réalité à niveaux multiples que nous connaissons pour les animaux d'élevage. Dans de nombreux cas, il s'agit là d'un vrai défi, du fait que les données à niveaux multiples ne conservent pas l'indépendance qu'on leur prête au sein de données de différents niveaux. Par exemple, pour les multi-omiques, les ensembles de données présentent une forte dimensionalité (grand nombre de caractéristiques pour un nombre relativement faible d'échantillons rendant l'inférence et la prédiction particulièrement difficiles, sauf si de grandes quantités de données sont disponibles ce qui est limité en raison des coûts d'analyse. De plus, élaborer ces algorithmes afin de gérer le flux massif de données et présenter des résultats en temps réel ne font que rendre le défi plus difficile à relever.

D'un autre côté, il est crucial d'évaluer la qualité prédictive des méthodes par apprentissage machine et les comparer avec les approches statistiques chaque fois que cela est possible<sup>83</sup>. L'apprentissage machine est particulièrement centré sur l'amélioration de la capacité prédictive et possède une tendance au surapprentissage qui peut le rendre plutôt inutile si les caractéristiques utilisées pour la formation changent dans le monde réel. De nouveaux domaines de recherche, tels que l'apprentissage machine probabiliste (Probabilistic Machine Learning), ont vu le jour à l'interface entre les deux domaines<sup>84</sup> et peuvent être explorés à l'avenir.

### **3.4. Elaborer des systèmes opérationnels afin de faciliter la collecte, l'intégration, l'analyse, la visualisation des données en temps réel et leur partage pour aider à la prise de décision**

Le but ultime est d'offrir des solutions et une assistance aux Services vétérinaires dans le processus de prise de décision, l'analyse doit être intégrée et simplifiée. Nous ne pouvons pas attendre de l'utilisateur final (vétérinaires, éleveurs, diagnosticiens, etc.) qu'il se préoccupe de méthodologies, mais il faut que nous fournissions des outils qui soient prêts à suivre, visualiser, évaluer et générer des alertes « interprétables » des risques en temps réel en changeant les scénarii et en facilitant la communication entre les différentes parties prenantes. Il est évident que ces outils doivent être adaptés aux besoins et aux capacités des parties prenantes, faciliter l'interprétation des résultats, assurer un avertissement précoce des utilisateurs face aux anomalies et proposer une série de mesures d'attaque et de lutte en fonction des conséquences sanitaires et économiques. Tout cela, faut-il l'espérer avec une faible courbe d'apprentissage pour permettre aux utilisateurs finaux d'utiliser ces outils de façon efficace.

Différentes initiatives ont été démarrées pour intégrer les sources de collecte des données et des méthodes de recherche individuelle dans des outils opérationnels. En Europe, le projet DECIDE (<https://decideproject.eu>) a pour objectif d'élaborer des outils d'aide à la décision guidés par des données pour les syndromes respiratoires et gastro-intestinaux et de les intégrer dans les systèmes de gestion des élevages existants<sup>85</sup>. Le projet prévoit de développer des outils d'aide à la décision afin de les intégrer dans les systèmes d'élevages existants. Des plateformes, telles que Disease Bioportal (<https://bioportal.ucdavis.edu>) de l'Université Davis de Californie, proposent déjà des moyens d'intégrer des structures de données complexes avec une visualisation et une analyse en temps réel pour aider à la prise de décision. Les producteurs se servant du Disease Bioportal peuvent visualiser et analyser différents flux de données (épreuves diagnostiques, informations génomiques, mouvements des animaux, documents de production, résistance aux antimicrobiens, etc.), ce qui relie leurs résultats de diagnostics moléculaires avec des informations au niveau des animaux et de la population dans cette plate-forme. Ces informations peuvent ensuite être converties en graphiques descriptifs, en visualisation géographique et en arbres phylogénétiques en temps réel pour faciliter l'évaluation des risques tout comme accélérer la communication relative aux risques entre les laboratoires, les vétérinaires et les producteurs. Le programme du BioPortal sur les maladies est actuellement en train d'ajouter l'apprentissage machine pour automatiser certaines analyses et faciliter le processus de génération de tableaux de bord, améliorant ainsi l'accès et l'expérience des utilisateurs finaux tout en incorporant de nouveaux algorithmes facilitant l'interprétation des données tout en mettant en lumière des événements tels que l'augmentation de l'incidence de la maladie, de nouvelles souches virales, des chutes de production, etc.

Le défi est toujours d'importance, néanmoins l'innovation et l'offre de méthodes sont plus faciles que jamais. Les nouvelles plateformes pouvant être partagées, telles que GitHub, qui permettent de partager les codes font progresser pour tous la mise en application des méthodes. Il existe également un logiciel d'application largement disponible permettant de faire des expérimentations en matière de création d'interfaces et de tableaux de bord pouvant être utiles pour les utilisateurs finaux, tels que Shiny, qui donne l'occasion de se servir de façon interactive d'une méthodologie fondée sur la recherche et examinée par des pairs au travers d'un réseau mondial d'utilisateurs. Par exemple, des tableaux de bord Shiny sont maintenant disponibles pour une interprétation radiographique, une visualisation d'omique, d'inférence de gène, d'analyse de réseau pour ne citer que quelques exemples<sup>86-90</sup>. Il est important de souligner que la plupart des tableaux de bord Shiny sont générés avec un logiciel en accès libre ce qui contribue au caractère ouvert du processus de numérisation bien qu'il y ait beaucoup de chemin à parcourir pour faire avancer la transition numérique des petits éleveurs et des Services vétérinaires dans les milieux à faible revenu<sup>91</sup>.

### 3.5. Former la prochaine génération des experts en données ainsi que les « travailleurs de santé animale/scientifiques en données vétérinaires »

La prochaine génération des travailleurs de santé animale doit avoir non seulement connaissance de la santé vétérinaire mais également des connaissances dans divers domaines qui aident au traitement et à l'analyse de données qui sont tout simplement trop importantes pour être utilisées avec des outils classiques. Les futurs épidémiologistes vétérinaires devront aussi connaître le traitement parallèle avec l'apprentissage machine, les systèmes de technologie de l'information et intégrer des compétences techniques, telles que la programmation par ordinateur qui ne fait pas traditionnellement partie de leur formation. Pour aborder ces questions, il y a globalement deux approches : adapter les formations existantes afin d'intégrer la connaissance nécessaire pour ces nouvelles exigences et constituer des équipes interdisciplinaires ayant différentes capacités et différentes formations. **L'élaboration d'une formation innovante** peut se faire en adaptant la formation existante apprentissage machine/science des données mais orientée vers la santé animale. Il faut donc que la formation touche les membres ayant connaissance des besoins vétérinaires mais aussi des experts en sciences des ordinateurs pouvant proposer des instructions solides, pratiques et accessibles en matière de techniques d'apprentissage machine, des boîtes à outils et des connaissances et ainsi, faire la jonction entre ces deux domaines. La participation de membres de différentes facultés disposant de divers domaines d'expertise permet d'offrir aux étudiants une formation bien construite<sup>92</sup>. A cet égard, différentes initiatives sont apparues ces dernières années afin d'aboutir à une science des données transdisciplinaire en ayant, par exemple, recours à des programmes de co-enseignement (<https://datascience.duke.edu/>). Une autre approche a été mise en œuvre à Cornell, un programme appelé Discipline-based Educational Researcher program, où un concepteur de programme pédagogique est intégré dans un département afin de contribuer à déterminer quelles sont les méthodes d'enseignement actives les plus appropriées pour ce sur quoi travaille le département (<https://cder.as.cornell.edu>). La coopération d'étudiants venant de différents domaines et de disciplines diverses pour se pencher sur la résolution de problèmes ouverts pratiques peut également contribuer à développer des compétences pour leur pratique future. Cette approche favorise les aptitudes de pensée critique, développe les capacités d'examen des documents existants et encourage un apprentissage en continu souvent dans le cadre d'un environnement

d'équipe. Des programmes similaires sont mis en œuvre dans des programmes de mastère spécialisés mais ils peuvent aussi être mis en œuvre dans la formation de doctorat en médecine vétérinaire.

**Mettre en place une équipe interdisciplinaire** travaillant en collaboration est essentiel pour le succès de la numérisation dans les Services vétérinaires et les professionnels doivent être préparés à travailler dans cet environnement. De bons principes et des meilleures pratiques ont été identifiés dans les documents publiés, y compris de bonnes communications, des réunions régulières, un mélange approprié de compétences et une expertise dans des domaines, une vision claire et des objectifs bien définis, une bonne direction et une bonne gestion, une culture de l'esprit d'équipe, une formation et des possibilités de développement, une flexibilité, etc.<sup>93</sup>

En plus des compétences techniques, la communication représente également un atout important pour les professionnels de l'épidémiologie vétérinaire de précision. Le succès dépend en partie de la façon dont les utilisateurs finaux vont se servir du produit, donc malgré la complexité de la collecte, de l'intégration, de la cartographie et de l'analyse des données, les résultats doivent être d'un accès, d'une compréhension et d'une utilisation faciles. Les experts en interaction êtres humains-ordinateurs sont essentiels dans ce processus pour orienter la conception de logiciels conviviaux et concevoir son utilisation par les utilisateurs finaux. D'un point de vue opérationnel, une co-conception ainsi qu'une implication des utilisateurs finaux sont essentiels pour garantir le succès de l'application et profiter de sa valeur. De la même façon, il est très important d'avoir des suivis visant à recevoir des retours d'information ou étudier et communiquer sur les avantages d'un point de vue pratique, économique et sanitaire présentant les objectifs mesurables et palpables.

## Conclusion

Nous pensons que la numérisation des Services vétérinaires peut considérablement améliorer la santé animale, la sécurité alimentaire et la sécurité sanitaire des aliments si le processus de numérisation devient équitable, accessible et éthique pour les parties prenantes de tout type et de toute taille. Elle peut aussi contribuer à augmenter de façon substantielle la durabilité, le bien-être animal et la rentabilité dans le fonctionnement de l'élevage. La numérisation exige une intégration en temps réel vers diverses sources de données à plusieurs échelles ainsi qu'un accès convivial et sûr et une visualisation de ce type d'informations par des utilisateurs finaux (à l'aide de tableaux de bord par exemple) afin de pouvoir donner en temps voulu des éléments pour la prise de décision. Nous croyons qu'en s'attaquant aux cinq défis existant en matière de numérisation pour la santé animale, il y aura une amélioration significative de la santé animale, du bien-être animal, de la productivité des exploitations et des revenus de la filière de l'élevage. Bien que l'accent soit mis dans le présent rapport sur le domaine de la santé des animaux d'élevage et la sécurité sanitaire des aliments, étant donné que ce sont les principaux domaines couverts par les services vétérinaires, nous sommes convaincus que la transformation numérique et l'utilisation de l'épidémiologie de précision représente une approche holistique, du type « Une seule santé » qui pourrait effectivement être utilisée pour mieux s'attaquer aux problèmes au niveau de la population dans tous les milieux de santé publique et de santé animale (par exemple en partant de l'étude de l'impact des expositions environnementales, des maladies infectieuses ou des maladies chroniques chez l'être humain pour aller jusqu'à l'évaluation des questions épidémiologiques touchant les animaux de compagnie ou la conservation de la faune sauvage. Les Autorités devraient définir des priorités pour élaborer des politiques, proposer des recommandations et investir dans la numérisation (y compris la normalisation des données et l'interconnectivité) parce que non seulement cela aiderait les services vétérinaires à faire mieux, plus vite et plus efficacement leur travail mais cela apporterait également de nombreux bénéfices sociaux et économiques.

## Références bibliographiques

1. Diebold, F. X. (2012). "On the Origin(s) and Development of the Term "Big Data.""
2. Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*, 15(2). <https://doi.org/10.1371/journal.pone.0228987>
3. Laney, D. (2001). 3-D data management: controlling data volume, velocity and variety. In *META Group Research Note, February 6*.
4. McCue, M. E., & McCoy, A. M. (2017). The scope of Big Data in one medicine: Unprecedented opportunities and challenges. *Frontiers in Veterinary Science*, 4(NOV), 1–23. <https://doi.org/10.3389/fvets.2017.00194>
5. VanderWaal, K., Morrison, R. B., Neuhauser, C., Vilalta, C., & Perez, A. M. (2017). Translating big data into smart data for veterinary epidemiology. *Frontiers in Veterinary Science*, 4(JUL), 1–7. <https://doi.org/10.3389/fvets.2017.00110>
6. Andreu-Perez, J., Poon, C. C. Y., Merrifield, R. D., Wong, S. T. C., & Yang, G. Z. (2015). Big Data for Health. *IEEE Journal of Biomedical and Health Informatics*, 19(4), 1193–1208. <https://doi.org/10.1109/JBHI.2015.2450362>
7. Nigania, J. (2018). *The 10 V's of Big Data*. <http://houseofbots.com/news-detail/2819-1-the-10-v%27s-of-big-data>
8. Shafer, T. (2017). *The 42 V's of Big Data and Data Science*. <https://www.elderresearch.com/blog/42-v-of-big-data>
9. Smith, D., Lyle, S., Berry, A., & Manning, N. (2015). *Internet of Animal Health Things (IoAHT) Opportunities and Challenges*.
10. McCue, M. E., & McCoy, A. M. (2019). Harnessing big data for equine health. *Equine Veterinary Journal*, 51(4), 429–432. <https://doi.org/10.1111/evj.13080>
11. Benjamin, M., & Yik, S. (2019). Precision livestock farming in swine welfare: A review for swine practitioners. *Animals*, 9(4), 1–21. <https://doi.org/10.3390/ani9040133>
12. Sassi, N. Ben, Averós, X., & Estevez, I. (2016). Technology and poultry welfare. *Animals*, 6(10), 1–21. <https://doi.org/10.3390/ani6100062>
13. Piñeiro, C., Morales, J., Rodríguez, M., Aparicio, M., Manzanilla, E. G., & Koketsu, Y. (2019). Big (pig) data and the internet of the swine things: A new paradigm in the industry. *Animal Frontiers*, 9(2), 6–15. <https://doi.org/10.1093/af/vfz002>
14. Knight, C. H. (2020). Review: Sensor techniques in ruminants: More than fitness trackers. *Animal*, 14(S1), S187–S195. <https://doi.org/10.1017/S1751731119003276>
15. Kleen, J. L., & Guatteo, R. (2023). Precision Livestock Farming: What Does It Contain and What Are the Perspectives? In *Animals* (Vol. 13, Issue 5). MDPI. <https://doi.org/10.3390/ani13050779>
16. Ladner, J. T., Grubaugh, N. D., Pybus, O. G., & Andersen, K. G. (2019). Precision epidemiology for infectious disease control. *Nature Medicine*, 25(2), 206–211. <https://doi.org/10.1038/s41591-019-0345-2>
17. Khoury, M. J., Iademarco, M. F., Riley, W. T., Sciences, P., Service, P. H., Corps, C., Precision, T., & Initiative, M. (2016). Precision Public Health for the Era of Precision Medicine - ClinicalKey. *Am J Prev Med*, 50(3), 398–401. <https://doi.org/10.1016/j.amepre.2015.08.031.Precision>
18. Van Borm, S., Belák, S., Freimanis, G., Fusaro, A., Granberg, F., Höper, D., King, D. P., Monne, I., Orton, R., & Rosseel, T. (2014). Next-generation sequencing in veterinary medicine: How can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases? *Methods in Molecular Biology*, 1247, 415–436. [https://doi.org/10.1007/978-1-4939-2004-4\\_30](https://doi.org/10.1007/978-1-4939-2004-4_30)

19. Bishop, S. C., & Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livestock Science*, 166(1), 190–198. <https://doi.org/10.1016/j.livsci.2014.04.034>
20. Islam, M. A., Rony, S. A., Rahman, M. B., Cinar, M. U., Villena, J., Uddin, M. J., & Kitazawa, H. (2020). Improvement of disease resistance in livestock: Application of immunogenomics and CRISPR/Cas9 technology. In *Animals* (Vol. 10, Issue 12, pp. 1–20). MDPI AG. <https://doi.org/10.3390/ani10122236>
21. Nietfeld, F., Höltig, D., Willems, H., Valentin-Weigand, P., Wurmser, C., Waldmann, K. H., Fries, R., & Reiner, G. (2020). Candidate genes and gene markers for the resistance to porcine pleuropneumonia. *Mammalian Genome*, 31(1–2), 54–67. <https://doi.org/10.1007/s00335-019-09825-0>
22. Lagumdzic, E., Pernold, C., Viano, M., Olgiati, S., Schmitt, M. W., Mair, K. H., & Saalmüller, A. (2022). Transcriptome Profiling of Porcine Naïve, Intermediate and Terminally Differentiated CD8+ T Cells. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.849922>
23. Pluta, A., Taxis, T. M., van der Meer, F., Shrestha, S., Qualley, D., Coussens, P., Rola-Łuszczak, M., Ryło, A., Sakhawat, A., Mamanova, S., & Kuźmak, J. (2023). An immunoinformatics study reveals a new BoLA-DR-restricted CD4+ T cell epitopes on the Gag protein of bovine leukemia virus. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-48899-4>
24. Bagnicka, E., Kawecka-Grochocka, E., Pawlina-Tyszko, K., Zalewska, M., Kapusta, A., Kościuczuk, E., Marczak, S., & Ząbek, T. (2021). MicroRNA expression profile in bovine mammary gland parenchyma infected by coagulase-positive or coagulase-negative staphylococci. *Veterinary Research*, 52(1). <https://doi.org/10.1186/s13567-021-00912-2>
25. Islam, M. A., Große-Brinkhaus, C., Pröll, M. J., Uddin, M. J., Rony, S. A., Tesfaye, D., Tholen, E., Hölker, M., Schellander, K., & Neuhoff, C. (2016). Deciphering transcriptome profiles of peripheral blood mononuclear cells in response to PRRSV vaccination in pigs. *BMC Genomics*, 17(1). <https://doi.org/10.1186/s12864-016-2849-1>
26. Kreitmaier, P., Katsoula, G., & Zeggini, E. (2023). Insights from multi-omics integration in complex disease primary tissues. In *Trends in Genetics* (Vol. 39, Issue 1, pp. 46–58). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2022.08.005>
27. Belák, S., Karlsson, O. E., Leijon, M., & Granberg, F. (2013). December 2013 in issue 32-3 of the Scientific and Technical Review. (1) Department of Virology. In *Rev. sci. tech. Off. int. Epiz* (Vol. 32, Issue 3).
28. Hernández-Triana, L. M., Garza-Hernández, J. A., Ortega Morales, A. I., Prosser, S. W. J., Hebert, P. D. N., Nikolova, N. I., Barrero, E., de Luna-Santillana, E. de J., González-Alvarez, V. H., Mendez-López, R., Chan-Chable, R. J., Fooks, A. R., & Rodríguez-Pérez, M. A. (2021). An Integrated Molecular Approach to Untangling Host–Vector–Pathogen Interactions in Mosquitoes (Diptera: Culicidae) From Sylvan Communities in Mexico. *Frontiers in Veterinary Science*, 7. <https://doi.org/10.3389/fvets.2020.564791>
29. Hickman, R. A., Leangapichart, T., Lunha, K., Jiwakanon, J., Angkititrakul, S., Magnusson, U., Sunde, M., & Järhult, J. D. (2021). Exploring the Antibiotic Resistance Burden in Livestock, Livestock Handlers and Their Non-Livestock Handling Contacts: A One Health Perspective. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.651461>
30. Oniciuc, E. A., Likotrafiti, E., Alvarez-Molina, A., Prieto, M., Santos, J. A., & Alvarez-Ordóñez, A. (2018). The present and future of whole genome sequencing (WGS) and whole metagenome sequencing (WMS) for surveillance of antimicrobial resistant microorganisms and antimicrobial resistance genes across the food chain. In *Genes* (Vol. 9, Issue 5). MDPI AG. <https://doi.org/10.3390/genes9050268>
31. Gaudette, S., Smart, L., Woodward, A. P., Sharp, C. R., Hughes, D., Bailey, S. R., Dandrieux, J. R. S., Santos, L., & Boller, M. (2023). Biomarkers of endothelial activation and inflammation in dogs with organ dysfunction secondary to sepsis. *Frontiers in Veterinary Science*, 10. <https://doi.org/10.3389/fvets.2023.1127099>
32. Perera, T. R. W., Skerrett-Byrne, D. A., Gibb, Z., Nixon, B., & Swegen, A. (2022). The Future of Biomarkers in Veterinary Medicine: Emerging Approaches and Associated Challenges. In *Animals* (Vol. 12, Issue 17). MDPI. <https://doi.org/10.3390/ani12172194>

33. Koene, M. G. J., Mulder, H. A., Stockhofe-Zurwieden, N., Kruijt, L., & Smits, M. A. (2012). Serum protein profiles as potential biomarkers for infectious disease status in pigs. *BMC Veterinary Research*, 8. <https://doi.org/10.1186/1746-6148-8-32>
34. Caja, G., Castro-Costa, A., & Knight, C. H. (2016). Engineering to support wellbeing of dairy animals. *Journal of Dairy Research*, 83(2), 136–147. <https://doi.org/10.1017/S0022029916000261>
35. Michie, C., Andonovic, I., Davison, C., Hamilton, A., Tachtatzis, C., Jonsson, N., Duthie, C. A., Bowen, J., & Gilroy, M. (2020). The Internet of Things enhancing animal welfare and farm operational efficiency. *Journal of Dairy Research*, 87(S1), 20–27. <https://doi.org/10.1017/S0022029920000680>
36. Blewey, J. (2010). Precision dairy farming: advanced analysis solutions for future profitability. *Proceedings of the First North American Conference on Precision Dairy Management*.
37. Connolly, A. (2019). The new Digital World of Dairy Farming - Bridging the data gap. *Proceedings of the II International Precision Dairy Farming Conference*, 1–8.
38. Crowe, M. A., Hostens, M., & Opsomer, G. (2018). Reproductive management in dairy cows - The future. *Irish Veterinary Journal*, 71(1), 1–13. <https://doi.org/10.1186/s13620-017-0112-y>
39. Hudson, C., Kaler, J., & Down, P. (2018). Use of big data in cattle practice. *In Practice*, 983.
40. Thorup, V. M., Nielsen, B. L., Robert, P. E., Giger-Reverdin, S., Konka, J., Michie, C., & Friggens, N. C. (2016). Lameness affects cow feeding but not rumination behavior as characterized from sensor data. *Frontiers in Veterinary Science*, 3(MAY), 1–11. <https://doi.org/10.3389/fvets.2016.00037>
41. Džermeikaitė, K., Bačėninaitė, D., & Antanaitis, R. (2023). Innovations in Cattle Farming: Application of Innovative Technologies and Sensors in the Diagnosis of Diseases. In *Animals* (Vol. 13, Issue 5). MDPI. <https://doi.org/10.3390/ani13050780>
42. Berckmans, D. (2017). General introduction to precision livestock farming. *Animal Frontiers*, 7(1), 6–11. <https://doi.org/10.2527/af.2017.0102>
43. Bustamante, E., Guijarro, E., García-Diego, F. J., Balasch, S., Hospitaler, A., & Torres, A. G. (2012). Multisensor system for isothermal measurements to assess indoor climatic conditions in poultry farms. *Sensors*, 12(5), 5752–5774. <https://doi.org/10.3390/s120505752>
44. da Fonseca, F. N., Abe, J. M., de Alencar Nääs, I., da Silva Cordeiro, A. F., do Amaral, F. V., & Ungaro, H. C. (2020). Automatic prediction of stress in piglets (*Sus Scrofa*) using infrared skin temperature. *Computers and Electronics in Agriculture*, 168. <https://doi.org/10.1016/j.compag.2019.105148>
45. Cornou, C., Lundbye-Christensen, S., & Kristensen, A. R. (2011). Modelling and monitoring sows' activity types in farrowing house using acceleration data. *Computers and Electronics in Agriculture*, 76(2), 316–324. <https://doi.org/10.1016/j.compag.2011.02.010>
46. Pandey, S., Kalwa, U., Kong, T., Guo, B., Gauger, P. C., Peters, D. J., & Yoon, K. J. (2021). Behavioral monitoring tool for pig farmers: Ear tag sensors, machine intelligence, and technology adoption roadmap. *Animals*, 11(9). <https://doi.org/10.3390/ani11092665>
47. Okada, H., Itoh, T., Suzuki, K., & Tsukamoto, K. (2009). Wireless sensor system for detection of avian influenza outbreak farms at an early stage. *Proceedings of the 2009 IEEE Sensors Conference*.
48. da Silva, J. P., de Alencar Nääs, I., Abe, J. M., & da Silva Cordeiro, A. F. (2019). Classification of piglet (*Sus Scrofa*) stress conditions using vocalization pattern and applying paraconsistent logic Et. *Computers and Electronics in Agriculture*, 166. <https://doi.org/10.1016/j.compag.2019.105020>
49. Besteiro, R., Arango, T., Rodríguez, M. R., Fernández, M. D., & Velo, R. (2018). Estimation of patterns in weaned piglets' activity using spectral analysis. *Biosystems Engineering*, 173, 85–92. <https://doi.org/10.1016/j.biosystemseng.2017.06.014>

50. Nasirahmadi, A., Richter, U., Hensel, O., Edwards, S., & Sturm, B. (2015). Using machine vision for investigation of changes in pig group lying patterns. *Computers and Electronics in Agriculture*, 119, 184–190. <https://doi.org/10.1016/j.compag.2015.10.023>
51. White, R. P., Schofield, C. P., Green, D. M., Parsons, D. J., & Whittemore, C. T. (2004). The effectiveness of a visual image analysis (VIA) system for monitoring the performance of growing/finishing pigs. *Animal Science*, 78(3), 409–418. [https://doi.org/DOI: 10.1017/S1357729800058811](https://doi.org/DOI:10.1017/S1357729800058811)
52. Guarino, M., Norton, T., Berckmans, D., Vranken, E., & Berckmans, D. (2017). A blueprint for developing and applying precision livestock farming tools: A key output of the EU-PLF project. *Animal Frontiers*, 7(1), 12–17. <https://doi.org/10.2527/af.2017.0103>
53. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. In *Agricultural Systems* (Vol. 153, pp. 69–80). Elsevier Ltd. <https://doi.org/10.1016/j.agsy.2017.01.023>
54. Trevisan G, Linhares LCM, Crim B, Dubey P, Schwartz KJ, Burrough ER, Main RG, Sundberg P, Thurn M, Lages PTF, Corzo CA, Torrison J, Henningson J, Herrman E, Hanzlicek GA, Raghavan R, Marthaler D, Greseth J, Clement T, Christopher-Hennings J, Linhares DCL. (2019). Macroepidemiological aspects of porcine reproductive and respiratory syndrome virus detection by major United States veterinary diagnostic laboratories over time, age group, and specimen. *PLoS One*. 14(10):e0223544. doi: 10.1371/journal.pone.0223544.
55. Martínez-López, B., Perez, A. M., & Sánchez-Vizcaíno, J. M. (2009). Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transboundary and Emerging Diseases*, 56(4), 109–120. <https://doi.org/10.1111/j.1865-1682.2009.01073.x>
56. O'Hara KC, Beltrán-Alcrudo D, Hovari M, Tabakovski B, Martínez-López B. (2022). Network analysis of live pig movements in North Macedonia: Pathways for disease spread. *Front Vet Sci.*; 9:922412. doi: 10.3389/fvets.2022.922412.
57. Gao, W., Jiang, X., Hu, Z., Wang, Q., Shi, Y., Tian, X., Qiao, M., Zhang, J., Li, Y., & Li, X. (2023). Epidemiological investigation, determination of related factors, and spatial-temporal cluster analysis of wild type pseudorabies virus seroprevalence in China during 2022. *Frontiers in Veterinary Science*, 10. <https://doi.org/10.3389/fvets.2023.1298434>
58. O'Hara, K. C., Pires, A. F. A., & Martínez-López, B. (2021). Evaluating the association between climatic factors and sheep condemnations in the United States using cluster analysis and spatio-temporal modeling. *Preventive Veterinary Medicine*, 191. <https://doi.org/10.1016/j.prevetmed.2021.105342>
59. Baron, J. N., Aznar, M. N., Monterubbianesi, M., & Martínez-López, B. (2020). Application of network analysis and cluster analysis for better prevention and control of swine diseases in Argentina. *PLoS ONE*, 15(6). <https://doi.org/10.1371/journal.pone.0234489>
60. Mazzucato M, Marchetti G, Barbuiani M, Mulatti P, Fornasiero D, Casarotto C, Scolamacchia F, Manca G, Ferrè N. (2023). An integrated system for the management of environmental data to support veterinary epidemiology. *Front Vet Sci*. 10:1069979. doi: 10.3389/fvets.2023.1069979.
61. Arsevska, E., Roche, M., Hendriks, P., Chavernac, D., Falala, S., Lancelot, R., & Dufour, B. (2016). Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. *Computers and Electronics in Agriculture*, 123, 104–115. <https://doi.org/10.1016/j.compag.2016.02.010>
62. Tizzani, M., Muñoz-Gómez, V., de Nardi, M., Paolotti, D., Muñoz, O., Ceschi, P., Viltrop, A., & Capua, I. (2021). Integrating digital and field surveillance as complementary efforts to manage epidemic diseases of livestock: African swine fever as a case study. *PLoS ONE*, 16(12 December). <https://doi.org/10.1371/journal.pone.0252972>
63. Kandula, S., & Shaman, J. (2019). Reappraising the utility of Google Flu Trends. *PLoS Computational Biology*, 15(8). <https://doi.org/10.1371/journal.pcbi.1007258>
64. Prieto, V. M., Matos, S., Álvarez, M., Cacheda, F., & Oliveira, J. L. (2014). Twitter: A good place to detect health conditions. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0086191>

65. Yousefinaghani, S., Dara, R., Poljak, Z., Bernardo, T. M., & Sharif, S. (2019). The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-54388-4>
66. Bahlo, C., Dahlhaus, P., Thompson, H., & Trotter, M. (2019). The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Computers and Electronics in Agriculture*, 156(August 2018), 459–466. <https://doi.org/10.1016/j.compag.2018.12.007>
67. Willette, M., Ponder, J. B., McRuer, D. L., & Clark, E. E. (2013). Wildlife health monitoring systems in North America: from sentinel species to public policy. In A. Aguirre, R. S. Ostfel, & P. Daszak (Eds.), *Conservation Medicine: Applied Cases of Ecological Health* (pp. 552–562). Oxford University Press.
68. Main, R. (2016). Supporting VDL messaging and data analysis through data standardization. *Allen D. Leman Swine Conference*.
69. Papst, F., Saukh, O., Romer, K., Grandl, F., Jakovljevic, I., Steininger, F., Mayerhofer, M., Duda, J., & Egger-Danner, C. (2019, October 22). Embracing Opportunities of livestock big data integration with privacy constraints. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3365871.3365900>
70. Chetty, A., & Blekhman, R. (2024). Multi-omic approaches for host-microbiome data integration. In *Gut Microbes* (Vol. 16, Issue 1). Taylor and Francis Ltd. <https://doi.org/10.1080/19490976.2023.2297860>
71. Fiocchi, C. (2023). Omics and Multi-Omics in IBD: No Integration, No Breakthroughs. In *International Journal of Molecular Sciences* (Vol. 24, Issue 19). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ijms241914912>
72. Picard, M., Scott-Boyer, M. P., Bodein, A., Périn, O., & Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. In *Computational and Structural Biotechnology Journal* (Vol. 19, pp. 3735–3746). Elsevier B.V. <https://doi.org/10.1016/j.csbj.2021.06.030>
73. Hu, R., Yao, R., Li, L., Xu, Y., Lei, B., Tang, G., Liang, H., Lei, Y., Li, C., Li, X., Liu, K., Wang, L., Zhang, Y., Wang, Y., Cui, Y., Dai, J., Ni, W., Zhou, P., Yu, B., & Hu, S. (2022). A database of animal metagenomes. *Scientific Data*, 9(1). <https://doi.org/10.1038/s41597-022-01444-w>
74. Shah, H. A., Liu, J., Yang, Z., & Feng, J. (2021). Review of Machine Learning Methods for the Prediction and Reconstruction of Metabolic Pathways. In *Frontiers in Molecular Biosciences* (Vol. 8). Frontiers Media S.A. <https://doi.org/10.3389/fmolb.2021.634141>
75. Grzesiak W and Zaborski D (2012). Examples of the Use of Data Mining Methods in Animal Breeding. Data Mining Applications in Engineering and Medicine. InTech. Available at: <http://dx.doi.org/10.5772/50893>.
76. Williams ML, Mac Parthaláin N, Brewer P, James WPJ, Rose MT. (2016). A novel behavioral model of the pasture-based dairy cow from GPS data using data mining and machine learning techniques. *J Dairy Sci*. 99(3):2063-2075. doi: 10.3168/jds.2015-10254.
77. Vahabi, N., & Michailidis, G. (2022). Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review. In *Frontiers in Genetics* (Vol. 13). Frontiers Media S.A. <https://doi.org/10.3389/fgene.2022.854752>
78. Faverjon, C., Bernstein, A., Grütter, R., Nathues, C., Nathues, H., Sarasua, C., Sterchi, M., Vargas, M. E., & Berezowski, J. (2019). A transdisciplinary approach supporting the implementation of a big data project in livestock production: An example from the swiss pig production industry. *Frontiers in Veterinary Science*, 6(JUL), 1–11. <https://doi.org/10.3389/fvets.2019.00215>
79. van der Weerd, C. A., Vonder, M. R., & Berkers, F. (2019). Accelerating adoption of precision dairy farming through digital innovation hubs. *Proceedings of the Second International Precision Dairy Farming Conference*, 8–11.
80. Herland, M., Khoshgoftaar, T. M., & Wald, R. (2014). A review of data mining using big data in health informatics. *Journal of Big Data*, 1, 2. <https://doi.org/10.1021/ci0255782>

81. Ezanno, P., Picault, S., Beaunée, G., Bailly, X., Muñoz, F., Duboz, R., Monod, H., & Guégan, J. F. (2021). Research perspectives on animal health in the era of artificial intelligence. In *Veterinary Research* (Vol. 52, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13567-021-00902-4>
82. Guitian, J., Arnold, M., Chang, Y., & Snary, E. L. (2023). Applications of machine learning in animal and veterinary public health surveillance. *OIE Revue Scientifique et Technique*, 42, 230–241. <https://doi.org/10.20506/RST.42.3366>
83. Hepworth, P. J., Nefedov, A. V., Muchnik, I. B., & Morgan, K. L. (2012). Broiler chickens can benefit from machine learning: Support vector machine analysis of observational epidemiological data. *Journal of the Royal Society Interface*, 9(73), 1934–1942. <https://doi.org/10.1098/rsif.2011.0852>
84. Ghahramani, Z. (2019). *Probabilistic modelling, machine learning, and the information revolution*. Citeseer.
85. van Schaik, G., Hostens, M., Faverjon, C., Jensen, D. B., Kristensen, A. R., Ezanno, P., Frössling, J., Dórea, F., Jensen, B.-B., Carmo, L. P., Steeneveld, W., Rushton, J., Gilbert, W., Bearth, A., Siegrist, M., Kaler, J., Ripperger, J., Siehler, J., de Wit, S., ... Nielen, M. (2023). The DECIDE project: from surveillance data to decision-support for farmers and veterinarians. *Open Research Europe*, 3, 82. <https://doi.org/10.12688/openreseurope.15988.1>
86. Badgeley, M. A., Liu, M., Glicksberg, B. S., Shervey, M., Zech, J., Shameer, K., Lehar, J., Oermann, E. K., McConnell, M. V., Snyder, T. M., & Dudley, J. T. (2019). CANDI: An R package and Shiny app for annotating radiographs and evaluating computer-aided diagnosis. *Bioinformatics*, 35(9), 1610–1612. <https://doi.org/10.1093/bioinformatics/bty855>
87. Class, C. A., Ha, M. J., Baladandayuthapani, V., & Do, K. A. (2018). IDINGO-integrative differential network analysis in genomics with Shiny application. *Bioinformatics*, 34(7), 1243–1245. <https://doi.org/10.1093/bioinformatics/btx750>
88. Fawcett, L. (2018). Using Interactive Shiny Applications to Facilitate Research-Informed Learning and Teaching. *Journal of Statistics Education*, 26(1), 2–16. <https://doi.org/10.1080/10691898.2018.1436999>
89. Schultheis, H., Kuenne, C., Preussner, J., Wiegandt, R., Fust, A., Bentsen, M., & Looso, M. (2019). WILSON: Web-based Interactive Omics VisualizatiON. *Bioinformatics*, 35(6), 1055–1057. <https://doi.org/10.1093/bioinformatics/bty711>
90. Vinogradova, S., Saksena, S. D., Ward, H. N., Vigneau, S., & Gimelbrant, A. A. (2019). MaGIC: A machine learning tool set and web application for monoallelic gene inference from chromatin. *BMC Bioinformatics*, 20(1), 1–5. <https://doi.org/10.1186/s12859-019-2679-7>
91. Lioutas E.D., Charatsari C., De Rosa, M., (2021). Digitalization of agriculture: A way to solve the food problem or a trolley dilemma?. *Technology in Society*, 67, 101744. <https://doi.org/10.1016/j.techsoc.2021.101744>
92. National Academies of Sciences Engineering. (2017). Data Science Education in the Future. In *Envisioning the data science discipline: the undergraduate perspective: interim report ahead* (pp. 19–25). The National Academic Press.
93. Nancarrow, S. A., Booth, A., Ariss, S., Smith, T., Enderby, P., & Roots, A. (2013). Ten principles of good interdisciplinary teamwork. *Human Resources for Health*, 11(1), 1. <https://doi.org/10.1186/1478-4491-11-19>