

Groupe de travail sur les données Annexe 11

Audition de Gilles Salvat, Nicolas Canivet, Christophe Cordevant, Anses, le 30/08/21

Pour l'Académie : Christophe Brard (ChB), Francis Desbrosse (FD), Christian Hervé (ChH), Jean-Pierre Jégou (JPJ).

Pour l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses) : Gilles Salvat (GS), Inspecteur Général de Santé Publique Vétérinaire, Directeur Général Délégué du Pôle Recherche et Référence, Nicolas Canivet (NC), Directeur de la Stratégie et des programmes, Christophe Cordevant (Ch C), Responsable Valorisation, Direction de la Stratégie et des programmes. Tous les trois ont été informés de ce que l'audition est enregistrée.

Partage des données : Sujet technique et politique : il faut rassurer les éleveurs dans le cadre de nos discussions actuelles relatives au suivi sanitaire permanent des élevages qui implique un accès aux données par le vétérinaire (ChB).

Propriété et accès aux données (GS) : Cela reste un problème, même s'il s'agit de données publiques. Les demandes d'associations de défense des consommateurs, qui souhaitent connaître des noms de marques suite à des travaux menés par l'Agence et publiés dans des rapports (ex : contrôle de pesticides dans les couches culottes pour bébés, suite saisine de la DGCCRF), sont de plus en plus fréquentes ; l'ANSES répond non à la raison que son rôle est seulement d'informer d'un risque. Le droit des affaires est également avancé pour la protection des données (exemple des dossiers d'AMM). Si on prend le cas des souches bactériennes en médecine humaine, elles sont la propriété du centre national de référence ; en vétérinaire cela n'est pas clair (NC). Il y a différents types de données : celles générées par l'activité de recherche et de référence de l'Anses, celles collectées par l'ANSES (ex : Resapath, Resumeq, Vigimyc, Salmonella, RNOEA), celles générées à partir de souches collectées par elle. Au plan européen, un plan de gestion des données est établi dès le lancement d'un projet, qui aborde la gestion et la mise à disposition des données (bien commun pour la connaissance scientifique).

Voir aussi le paragraphe sur les Chartes d'Adhésion et d'Utilisation (GS).

Charte d'adhésion au réseau de l'Anses (GS), document contractuel signé par l'organisme propriétaire des données, qui accepte le partage de ses données ; c'est le cas pour les enquêtes épidémiologiques sur l'Antibiorésistance (remontée des antibiogrammes par les laboratoires d'analyses) avec le réseau RESAPATH ; c'est le cas également pour les prélèvements dans le cadre de travaux de recherche. Dans cette charte, l'ANSES s'engage à l'anonymisation, à préciser l'usage pour lequel la donnée a été collectée (s'il devait-être

modifié au cours d'une étude, un consentement éclairé serait nécessaire), parfois au stockage limité dans le temps, au respect des directives de la CNIL lorsque le propriétaire de la donnée est une personne physique. Lorsque le fournisseur de la donnée est un éleveur, l'ANSES se contente d'une information seule.

Stockage des données : (GS, NC et ChC). Le volume des données augmente. Une réflexion (NC) est engagée sur la gestion et la valorisation des données avec un impact sur divers domaines :

- Serveurs internes et externes (de stockage, de calculs) avec une nouvelle approche technologique, celle des Big Data.

- Sécurité contre le piratage (ChB) ? Re : c'est un risque difficile à maîtriser ; duplicata (vs la perte des données) et réseau de secours (vs blocage du système) ne sont pas forcément efficaces. L'ANSES n'est pas fortement exposée car les données qu'elle stocke sont destinées à être divulguées.

- Encadrement nécessaire des données (GS), notions de données « sensibles » et de secret professionnel. Exemple des autocontrôles par les industriels de la chaîne alimentaire dans le cadre de la Santé Publique Vétérinaire (SPV), qui génèrent des données confidentielles qui restent dans les usines, mais qui ne fournissent que des données brutes sur l'identification de l'espèce de bactérie sans détailler la souche ni la pathogénicité, d'où des discussions difficiles pour signer la charte. La notion de SPV permet parfois à l'ANSES, dans le cadre d'une enquête, de transgresser la charte : c'est le cas dans les plans de contrôle menés par les ISPV, afin de prendre les mesures adéquates.

- Mise en forme des données. Chaque usine fournit des millions de données (GS) qu'il faut mettre en forme, un algorithme de tri est à créer.

- L'ANSES a une stratégie pour constituer un patrimoine (NC) de données de Laboratoire qui sert à la communauté dans le cadre de sa politique de valorisation des données.

Qualité des données : La qualité des données est à vérifier (NC). L'Anses procède de façon mixte, d'une part "manuellement" par démarche expert, en fonction de la renommée du producteur de la donnée en y associant d'autres critères qui sont fonction du type de données et de la façon dont on les utilise ; par exemple, les données publiques de séquençage peuvent être identifiées sous divers noms de genre et d'expert, il faut alors creuser et voir qui est le déposant initial, d'autre part à l'aide de petites applications informatiques spécifiques à certains domaines qui permettent à des labos de vérifier en temps réel des anomalies. Une application "Qualiplan" adaptée à la chaîne alimentaire est en cours d'élaboration depuis 3 ans. Cependant, il apparaît que dans l'état actuel il est impossible de qualifier une donnée individuelle (GS) ; c'est là qu'intervient la notion de la correction du manque de qualité par la quantité massive, par exemple dans le cas des données participatives : c'est le cas de l'application "CITIQUE" qui permet le signalement géo-localisé des piqûres de tiques. Ceci nécessite un traitement différent des données participatives par rapport aux données professionnelles.

Cas des objets connectés (GS) : En élevage et chez le véto ils permettent le diagnostic "au cul de la vache", y compris pour les maladies à déclaration obligatoire. Le déploiement des données, via les smartphones munis d'applications, se fait en temps réel et remonte de plus en plus facilement avec un encadrement de plus en plus difficile, ce qui expose au risque de détournement malveillant. Ce fut le cas pour les végétaux où une marque américaine équipait ses moissonneuses batteuses de dispositifs de mesure du degré d'humidité du blé, ce qui lui

permettait d'avoir la donnée 72H avant le cours officiel et de pratiquer le délit d'initié à la bourse de Chicago en spéculant aux dépens des agriculteurs. L'encadrement est d'autant plus difficile qu'il dépend du type de données et de la façon dont on les utilise.

Partage des Données internes et externes (NC) : Cela fait partie d'une nouvelle stratégie, comment diffuser de façon contrôlée (séquence vs génome complet). L'Europe préconise le concept d'Open Data, mais elle a tendance à faire machine arrière car les USA ne jouent pas le jeu. Ce qui est financé par l'Europe doit permettre un retour. Il y a le cas des données financées par l'Europe exploitées par des industriels sans retour. Voir aussi le paragraphe sur les Chartes d'Adhésion et d'Utilisation (GS)

Déstockage : Question de ChB, Re GS : c'est un problème, actuellement l'ANSES ne déstocke pas. Un biologiste a du mal à déstocker. Pour les versions papier des dossiers d'AMM c'est 40 ans et 10 ans pour le reste. Re NC : pour limiter le stockage de données brutes on peut avoir intérêt à conserver physiquement une souche car le stockage informatique du séquençage d'une bactérie peut occuper 10 à 100 giga octets avec un coût financier et environnemental non négligeable. À l'avenir l'ADN synthétique serait une donnée plus compacte. ChC : on peut déstocker par choix, mais cela conditionne la recherche de demain. L'institut Max Planck a pu ainsi montrer la coévolution des salmonelles avec les animaux de la préhistoire à nos jours. En France nous possédons, à l'institut Pasteur, la collection Léon Leminor unique au monde, recueil sur les bactéries et leur environnement extrait de cahiers de labo.