

Les conditions de la libre circulation des données agricoles

Par Sébastien PICARDAT
Directeur général d'Agdatahub

On estime que 2,5 milliards de milliards de données agricoles sont produites chaque jour à l'échelle mondiale. Sachant que l'analyse de ces données peut, par exemple, donner en temps réel le rendement d'un céréalier français en période de moisson, il est crucial d'encadrer leur circulation. C'est la voie qui est prise en Europe où l'agriculture est le deuxième secteur économique le plus avancé quant à la gestion de ses données, derrière l'industrie automobile. En effet, s'il est essentiel que ces données circulent, car elles sont importantes pour alimenter de nouveaux modèles d'intelligence artificielle et de nouveaux outils d'aide à la décision, cela doit se faire dans un cadre sécurisé et prendre en compte le consentement des agriculteurs à l'usage de leurs données. Le nouveau règlement européen "Data Governance Act" et le futur règlement européen "Data Act"¹ (pendant du RGPD pour les données non personnelles) intègrent d'ailleurs cette dimension.

En France, comme ailleurs dans le monde, les exploitations agricoles sont de plus en plus connectées : robots de traite, capteurs de vêlage ou météo, tracteurs pilotés par GPS. De l'élevage aux grandes cultures, toutes les activités agricoles sont concernées et touchées, quelle que soit la taille de l'exploitation, aussi bien les fermes dites conventionnelles que les fermes converties à l'agriculture biologique. La France compte plus de 380 000 exploitations agricoles connectées à plus de 85 000 partenaires et fournisseurs de services avec lesquels elles échangent tous types de données agricoles : des données géographiques, comme des données parcellaires, des données techniques relatives aux pratiques agricoles, ou encore des données environnementales (qui remontent des capteurs météo, par exemple). Le terme de *big data* pour le secteur agricole est donc complètement adapté et adopté.

Or, ces données, qui font partie des biens immatériels des exploitants agricoles, ont de la valeur à plusieurs niveaux. Elles sont essentielles pour améliorer les performances économiques de la « ferme France », mais aussi pour rendre notre agriculture plus durable et améliorer les conditions de production aussi bien que les conditions de vie des agriculteurs et le bien-être animal. Il faut toutefois pour cela que ces données agricoles circulent de manière fluide entre les acteurs de l'écosystème « AgriTech » (se dit des acteurs qui mettent au point et proposent des services digitaux pour le secteur agricole), dans le respect du consentement des agriculteurs à l'usage de leurs données.

L'État français s'est assez tôt penché sur la question de cette gouvernance des données, sous l'impulsion de Stéphane Le Foll, alors ministre de l'Agriculture. En 2015, avec le ministère de la Recherche, des personnalités du secteur agricole ont été missionnées afin de poser les bases d'un plan « Agriculture – Innovation 2025 ». Ce travail a fait apparaître l'hétérogénéité et la dispersion des *data* agricoles, mais aussi leur potentiel si elles pouvaient être interconnectées.

¹ https://ec.europa.eu/commission/presscorner/detail/fr/ip_22_1113

UNE PLATEFORME D'ÉCHANGE DE DONNÉES

À la suite de leurs recommandations, Jean-Marc Bournigal, alors président de l'Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (Irstea), s'est vu confier une nouvelle mission : préparer la mise en place opérationnelle d'un portail de données agricoles et proposer un plan d'action pour favoriser l'émergence de nouveaux services à partir de ces données. De là est née la plateforme d'échange de données agricoles API-Agro, alors portée par les Instituts techniques agricoles et les chambres d'agriculture, et aujourd'hui opérée par Agdatahub (entreprise fondée en 2020 dont les actionnaires sont la *holding* agricole API-Agro, constituée d'actionnaires publics et privés représentatifs du secteur, et la Banque des Territoires, groupe Caisse des Dépôts).

UN GUIDE EUROPÉEN DE BONNE CONDUITE

En parallèle, d'autres acteurs se sont mobilisés, au niveau français comme européen, pour aborder la question d'un usage maîtrisé des données agricoles. En 2018, des syndicats agricoles européens réunis autour du COPA-COGECA (Comité des organisations professionnelles agricoles de l'Union européenne-Confédération générale des coopératives agricoles) ont publié un code de conduite² relatif à l'usage des données collectées auprès des agriculteurs et des organisations professionnelles agricoles. Dans la foulée, les Jeunes Agriculteurs et la FNSEA (Fédération nationale des syndicats d'exploitants agricoles) ont créé le label et la charte Data-Agri³, soit treize principes à respecter pour les fournisseurs de *e-services* agricoles (applications, logiciels...) et concernant notamment la portabilité des données et le respect du consentement. Le code européen de bonne conduite et sa déclinaison française ont pour effet de clarifier les relations contractuelles entre les agriculteurs et leurs fournisseurs dans une logique de protection quant à l'accès aux données. Le cœur de ces deux documents est représentatif d'une vision agricole qui met l'exploitant au centre du secteur, à l'inverse de la vision allemande qui, elle, priorise le machinisme agricole et les agrofournisseurs (gros collecteurs de données).

Les documents du COPA-COGECA et de Data-Agri sont fondateurs et essentiels pour créer de la confiance et maîtriser l'usage des données agricoles, même s'ils sont encore peu appliqués par les acteurs de l'AgriTech. Alors que tous s'accordent sur la nécessité d'accéder à des données actuelles et fiables afin de proposer des services innovants, force est de constater que les agriculteurs et les éleveurs ne sont encore que trop peu consultés quand, par exemple, les données collectées sur leurs exploitations sont partagées ou même vendues à un tiers... Pour l'Union européenne, le "Data Governance Act" (DGA) et le futur "Data Act" sont sur le point de changer la donne, en rééquilibrant le rapport de force entre les agriculteurs et les fournisseurs de services digitaux.

UN RÈGLEMENT EUROPÉEN AMBITIEUX

Après le Règlement général pour les données personnelles (RGPD, publié au Parlement européen en 2016), qui a permis de sensibiliser le grand public et les entreprises à la valeur des données, l'Union européenne va réglementer les données non personnelles. Une grande avancée puisque les données non personnelles représentent à peu près 80 % des données échangées quotidiennement dans le monde. Le futur règlement européen sur les données doit permettre de libérer les données en vue de leur utilisation, mais dans

² https://www.cema-agri.org/images/publications/brochures/EU_Code_of_conduct_on_agricultural_data_sharing_by_contractual_agreement_2020_FRENCH.pdf

³ <https://www.data-agri.fr/>

le respect de règles qui se veulent équitables, pratiques et claires. Celles-ci vont ainsi déterminer simplement qui peut accéder à quelles données et les utiliser, dans quel but et pour combien de temps. En résumé, ce règlement européen précise qui peut créer de la valeur à partir des données. Comme tout secteur économique, l'agriculture va y être soumise. Afin de faire respecter l'obligation de consentement des agriculteurs avant tout échange de données, un routeur de consentements a été élaboré par Agdatahub. Il est basé sur les résultats du Projet Multipass porté par les Instituts techniques agricoles et financé par le CASDAR (Compte d'affectation spécial au développement agricole et rural). Il est interconnectable librement avec tous les gestionnaires de consentements du marché.

Ce "Data Act" va ainsi offrir de nouvelles possibilités d'utiliser des services reposant sur l'accès à ces données. Par exemple, aujourd'hui, un agriculteur qui dispose d'équipements de différentes marques est dépendant de chaque fabricant pour l'analyse de ses données. Une fois le futur règlement en vigueur, il pourra recevoir des conseils personnalisés de la part d'une entreprise collectant les données des différents équipements.

UN DATA SPACE SECTORIEL

Cette future réglementation sur les données non personnelles s'inscrit dans la stratégie européenne pour les données (2020), qui vise à faire de l'Union européenne un *leader* dans une société fondée sur les données, au travers d'un marché unique des données. L'ambition de la Commission européenne est la création de dix espaces européens communs des données, dont un *data space* Agriculture. Cela va permettre la libre circulation des données agricoles au sein de l'Union européenne, dans le respect des règles d'accès aux données et d'utilisation.

Cette ambition européenne est partagée au sein de l'association Gaia-X. Au départ initiative franco-allemande promouvant un *cloud* européen souverain, Gaia-X veut contribuer à l'émergence de *data spaces* sectoriels européens, bâtis autour de valeurs de confiance, de sécurité, de souveraineté et d'interopérabilité. L'agriculture figure dans les sept secteurs économiques prioritaires décidés par l'association, et l'entreprise Agdatahub (qui opère des plateformes de gestion du consentement et d'échange de données agricoles) a été sélectionnée comme l'un des projets phares de Gaia-X dans le cadre du déploiement du *data space* Agriculture au niveau européen.

Pour autant, si les données vont pouvoir circuler librement et de manière équitable pour les agriculteurs dans le respect de leur consentement, la question de leur portabilité se pose encore. En effet, les données agricoles ne sont pas standardisées. C'est pourquoi, en France, l'association Numagri veut pallier ce problème en animant des cas d'usages par filières agricoles en vue de créer un « langage commun » des données agricoles, de l'amont agricole à la première commercialisation, en passant par la production. L'objectif final étant de favoriser la création de valeur par le développement de cas d'usages répondant aux attentes des acteurs de l'écosystème et des consommateurs.

Consentement, plateformes d'échange, souveraineté et standardisation : en France et en Europe, les éléments nécessaires à la circulation sécurisée des données sont identifiés et en train de se mettre en place. Les prochains mois seront déterminants pour le secteur agricole.

Intelligence artificielle et assurance : vers un renversement du déséquilibre informationnel ?

Par Xavier VAMPARYS

Laboratoire OpAIE de Télécom Paris – Institut polytechnique de Paris

Grâce à l'intelligence artificielle, les assureurs deviennent capables d'inférer des informations cachées à partir des données obtenues, directement ou indirectement, de leurs assurés. L'asymétrie d'information, qui penchait essentiellement en faveur de l'assuré, détenteur d'informations sur sa situation, son état de santé, etc., non partagées avec l'assureur, est peut-être en train de s'inverser. C'est désormais l'assureur qui dispose potentiellement de plus d'informations sur l'assuré, et donc sur ses risques, que l'assuré lui-même. Ce renversement du déséquilibre informationnel n'est pas sans poser la question du statut des informations inférées, et notamment du droit d'accès des assurés à ces informations.

L'asymétrie d'information qui caractérise le lien entre un assuré et un assureur affecte négativement le marché de l'assurance. Certes, cette asymétrie n'est pas propre à l'assurance : de façon générale, un vendeur en sait plus que l'acheteur sur le produit ou service vendu, alors qu'en retour l'acheteur en sait plus que le vendeur sur le prix qu'il est disposé à payer. Mais cette asymétrie est particulièrement prégnante en assurance et a pour conséquence deux « inefficacités ».

- Tout d'abord, l'anti-sélection (ou sélection adverse) qui fait qu'un « côté du marché ne peut pas observer [...] la qualité des biens situés de l'autre côté » (Varian, 2014) : s'assureront ceux qui ont le plus intérêt à l'être, en d'autres termes, les « mauvais risques » que l'assureur ne sait *a priori* pas identifier¹. À terme, c'est le marché qui est menacé de disparition puisque les « mauvais risques » chassent les « bons risques », qui ne considèrent plus en avoir pour leur argent à un niveau de prime trop élevé pour couvrir leurs risques (Akerlof, 1970).
- Ensuite, l'aléa moral engendré par le fait qu'une des parties au contrat (l'assureur) ne peut pas contrôler l'action de l'autre partie (l'assuré) qui, une fois couverte, peut adopter des comportements à risque dont les conséquences seront supportées par l'assureur et *in fine* par les autres assurés. S'ajoutent à ces comportements à risque les pratiques frauduleuses de certains assurés, qui peuvent être tentés d'obtenir réparation de dommages qu'ils n'ont pas subis.

¹ Cette analyse classique est remise en cause par des études empiriques qui mettent en évidence l'existence d'une « sélection avantageuse » dans certains marchés d'assurance : ce ne sont pas les « mauvais risques » qui s'assurent le plus, mais davantage les personnes ayant la plus forte aversion au risque, personnes qui se comportent de façon plus prudente, rendant moins probable la survenance du sinistre contre lequel elles s'assurent. Pour une critique de l'anti-sélection, voir par exemple Siegelman, 2004.

Dans le cas de la sélection adverse comme celui de l'aléa moral, ce qui fait défaut, c'est une information complète, sur l'assuré dans le premier et sur son comportement dans le second. Cette information est soit impossible à obtenir, notamment pour des raisons réglementaires, soit trop onéreuse. Bien sûr, l'assureur ne reste pas sans réponse. Il peut ainsi prévoir un délai de carence pendant lequel l'assuré ne sera pas couvert (ce qui l'empêchera notamment de déclarer des sinistres certains ou déjà intervenus à la date de souscription de son contrat), des franchises ou des plafonds (qui font supporter à l'assuré une partie des sinistres), un système de bonus/malus (qui incite l'assuré à un comportement responsable) (Zajdenweber, 2006), voire même observer le comportement de l'assuré *via* des objets connectés, en matière de risque santé ou automobile par exemple. Mais l'assureur peut au mieux « approximer » l'information détenue par l'assuré.

L'intelligence artificielle (IA) vient perturber ce déséquilibre informationnel. L'IA se définit comme une technologie capable, à partir d'observations, d'inférer des règles ou informations non connues ou évidentes pour des humains. Autrement formulé, l'IA est une technologie de prédiction qui réduit l'incertitude (Agrawal *et al.*, 2018). Les applications concrètes de cette capacité prédictive ne manquent pas : en matière de santé par exemple, certaines maladies sont rendues plus aisément détectables par des systèmes d'IA. Autre illustration : en matière automobile, ce sont bientôt des véhicules parfaitement autonomes et moins accidentogènes (grâce notamment à leur capacité à prédire et prévenir les accidents) qui circuleront sur nos routes. Ceci n'est pas sans conséquence sur le rapport entre assureur et assuré. Grâce aux outils statistiques, l'assureur est aujourd'hui capable d'extraire de la masse de données lui étant accessibles des informations qui lui échappaient pour tarifier plus justement un risque ou en prévoir la sinistralité. L'assuré n'est de son côté pas totalement démuné pour traiter certaines de ces informations puisqu'il peut avoir recours à des courtiers ou des comparateurs, qui confirmeront la pertinence du choix d'une couverture. Il peut par ailleurs avoir accès à certaines données ayant une incidence sur cette couverture (par exemple une prédisposition à une maladie ou une indication sur la suffisance de son activité physique pour la prévenir) par une source utilisant des systèmes d'IA autre qu'un assureur.

Mais la balance des informations dévoilées par l'IA penche en faveur de l'assureur, si bien que l'asymétrie d'information semble se réduire, ce qui est *a priori* bénéfique pour l'assureur, l'assuré et le marché de l'assurance. Tout d'abord, l'IA permettra de diminuer le coût d'acquisition des données pertinentes pour l'assureur et facilitera la relation entre assureur et assuré, souvent rebuté par les questions qui lui sont posées. Ces données pourront désormais pour partie être inférées grâce à l'IA. L'« expérience client » des assurés en sera améliorée sans que l'assureur perde en données utiles. Ensuite, l'IA fera passer l'assurance d'une logique de réparation à une logique de prévention. Grâce à sa capacité à anticiper le sinistre et à inciter l'assuré à des pratiques (de santé, d'exercice physique, de conduite automobile, etc.) plus vertueuses, l'assureur pourra limiter la fréquence et l'ampleur des sinistres. Ainsi, c'est le bien-être individuel mais aussi collectif qui sera augmenté. Toutefois, cela suppose que l'assureur fasse sa mue du rôle de porteur de risques à celui de prestataire de services de prévention. À ce stade, les assureurs peinent à convaincre leurs assurés de l'intérêt de ces services et de la justesse de leur prix. On peut ensuite s'interroger sur la légitimité d'un assureur à prescrire à des individus l'adoption de tel ou tel comportement, voire à les pénaliser s'ils ne se conforment pas à certaines pratiques jugées « bonnes ». La Déclaration de Montréal, qui fait figure de référence en matière d'IA éthique et responsable, considère par exemple que l'utilisation de l'IA ne doit « pas restreindre l'éventail des choix de vie et des expériences personnelles »². Pour d'autres chartes éthiques, c'est l'autonomie de l'individu qui ne doit pas être remise en cause par l'IA. Enfin, si la prévention coûte moins cher que l'indemnisation, le coût de l'assurance diminuera, ce qui la rendra plus accessible aux personnes qui aujourd'hui

² Déclaration de Montréal pour un développement responsable de l'IA, 2018.

s'en dispensent faute de pouvoir en payer le prix. La réduction de certaines externalités négatives, comme la fraude, contribuera également à faire baisser ce coût.

Peut-on aller un cran plus loin et imaginer que l'IA et l'accès accru à diverses données (de santé ou comportementales par exemple) vont engendrer un renversement de l'asymétrie d'information ? L'assureur en saurait alors plus que l'assuré sur la santé de ce dernier, son comportement et ses conséquences, ses goûts, préférences ou inclinations, voire ses désirs (une intention d'achat par exemple) avant même que celui-ci ne les ressente ou manifeste³. Ceci n'est pas sans poser question sur l'accès et l'usage qui sera fait par l'assureur de ces données personnelles, voire sensibles, inférées grâce à l'IA. Être capable d'identifier la dépression d'un individu à partir des messages qu'il laisse sur divers réseaux sociaux pour lui conseiller de consulter un psychologue est une chose, le faire pour lui vendre des cours de yoga ou un traitement en est une autre.

La question de l'usage de ces données de « sortie » (par opposition aux données d'« entrée » fournies par l'assuré) est d'abord celle de leur statut et des droits dont les assurés disposent sur celles-ci. Wachter *et al.* (2019) considèrent, après une revue des dispositions du Règlement général sur la protection des données (RGPD)⁴, des conclusions du groupe de travail « article 29 »^{5,6} et de la jurisprudence de la Cour de justice de l'Union européenne, sources de droit bien divergentes sur le sujet, que ces données inférées sont, au mieux, des données personnelles de « second rang » (*“economy class”*). Le RGPD ne semble par exemple conférer aux personnes concernées aucun droit sur ces données inférées, dont l'existence même n'a pas à leur être divulguée⁷. En d'autres termes, si un assuré fournit des informations pour obtenir une couverture par un assureur, il en connaîtra le tarif, mais pas les données inférées sous-jacentes⁸. Se créera alors une asymétrie d'information favorable à l'assureur, qui sera seul détenteur d'analyses ou prédictions pourtant potentiellement utiles à l'assuré. Le projet de règlement européen sur l'intelligence artificielle du 21 avril 2021 n'est pas plus ambitieux en ce domaine. Les personnes physiques faisant l'objet d'un traitement algorithmique ne sont informées⁹ d'un tel traitement que lorsqu'elles « interagissent » avec un système d'IA¹⁰ ou lorsque ce système reconnaît leurs émotions ou caractéristiques. On peut craindre qu'en raison du caractère intrusif des données de « sortie » et de l'accès à ce jour limité de l'assuré à ces données et, partant, de sa faculté à les contester, l'assuré cherche à restreindre les informations qu'il transmet, directement ou indirectement, à l'assureur. Cette hypothèse ne lui serait pas nécessairement favorable, l'absence de transparence étant généralement traitée avec suspicion par celui qui est en charge de son traitement, ici l'assureur.

³ Dans d'autres domaines, certains fournisseurs comme Amazon envisagent, grâce à l'IA, de livrer des produits avant même que leurs clients en aient fait la demande.

⁴ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

⁵ Le groupe de travail de l'article 29 traitait les questions relatives à la protection des données personnelles avant l'entrée en vigueur du RGPD. Il est aujourd'hui remplacé par le Comité européen de la protection des données (EDPB).

⁶ Le groupe de travail de l'article 29 considère que les inférences peuvent être des données personnelles. Wachter *et al.*, 2020, pp. 25-28.

⁷ Wachter *et al.*, 2020, relèvent toutefois que si les données n'ont pas été collectées auprès de la personne concernée, l'article 14 du RGPD ouvre potentiellement à cette personne le droit d'en être informée par le responsable de traitement.

⁸ Ces informations peuvent par ailleurs être protégées par le droit de la propriété intellectuelle et le droit au secret, ce qui constituerait un obstacle de plus pour l'accès de ces personnes aux informations inférées.

⁹ Leurs droits se limitent à cette simple information.

¹⁰ Article 52 du projet de règlement sur l'intelligence artificielle en date du 21 avril 2021.

La disposition des assurés à fournir plus de données à leur assureur en échange de produits ou services plus personnalisés (y compris s'agissant de la tarification du risque) est une question récurrente. Elle fait l'objet de nombre de sondages aux conclusions discordantes. Elle masque celle, peut-être plus essentielle, de l'accès des personnes concernées à l'information générée par l'IA. Cette question s'inscrit d'ailleurs dans une interrogation plus globale sur le partage d'informations personnelles entre un assuré et un assureur, voire entre un assureur et un tiers au contrat d'assurance¹¹, dans un contexte où le secteur de l'assurance s'engage sur deux voies diamétralement opposées : celle d'une accélération de la connaissance du risque des assurés (notamment permise par l'IA) et celle d'une réglementation qui referme progressivement le champ des informations pouvant être prises en compte par l'assureur. Le droit à l'oubli, qui permet depuis 2015 aux personnes ayant été affectées par certaines pathologies de ne pas les mentionner dans le questionnaire de santé auquel les assureurs les soumettent, et l'interdiction dudit questionnaire pour certains produits d'assurance-emprunteur depuis une loi du 28 février 2022¹² sont deux illustrations parlantes de ce renfermement.

BIBLIOGRAPHIE

AGRAWAL A., GANS J. & GOLDFARB A. (2018), *Prediction Machines – The Simple Economics of Artificial Intelligence*, Harvard Business Review Press.

AKERLOF G. (1970), "The market for lemons: Quality uncertainty and the market mechanism", *The Quarterly Journal of Economics*, 84(3), pp. 488-500.

SIEGELMAN P. (2004), "Adverse selection in insurance markets: An exaggerated threat", *The Yale Law Journal*, 113, pp. 1223-1281.

VARIAN H. (2014), *Introduction à la microéconomie*, De Boeck, 9^e édition.

WACHTER S. & MITTELSTADT B. (2019), "A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI", *Columbia Business Law Review*, 1.

ZADJENWEBER D. (2006), *Économie et Gestion de l'Assurance*, Économica.

¹¹ La Commission européenne propose l'adoption d'un cadre en matière de « finance ouverte », qui comprendrait notamment le partage d'informations entre établissements financiers mais également entre ces établissements et des tiers, dans le droit fil de ce qu'a accompli la directive sur les services de paiement (directive (UE) 2015/2366 du Parlement européen et du Conseil du 25 novembre 2015 concernant les services de paiement dans le marché intérieur). Voir « Communication de la Commission sur une stratégie en matière de finance numérique pour l'UE » en date du 24 septembre 2020 et "Commission staff working document on Common European Data Spaces" du 23 février 2022.

¹² Loi n°2022-270 du 28 février 2022.