

Les données, un enjeu stratégique

L'analyse de données est l'un des besoins les plus fréquemment exprimés lors des enquêtes IA menées auprès des coopératives agricoles. Pour qu'une analyse de données soit efficace, il est primordial de définir une stratégie claire : quels objectifs atteindre ? quels indicateurs suivre ? comment mesurer les retours sur investissement ?

- [Aspects quantitatifs, qualitatifs, sécuritaires](#)
- [Calibrer les données, un premier enjeu](#)
- [Usage des données](#)
- [Collecter les données](#)
- [Transformer les données en informations](#)

Aspects quantitatifs, qualitatifs, sécuritaires

La donnée, pour être utile, doit être gouvernée, qualifiée et intégrée dans un tableau de bord pertinent. Sa qualité est un facteur clé de réussite.

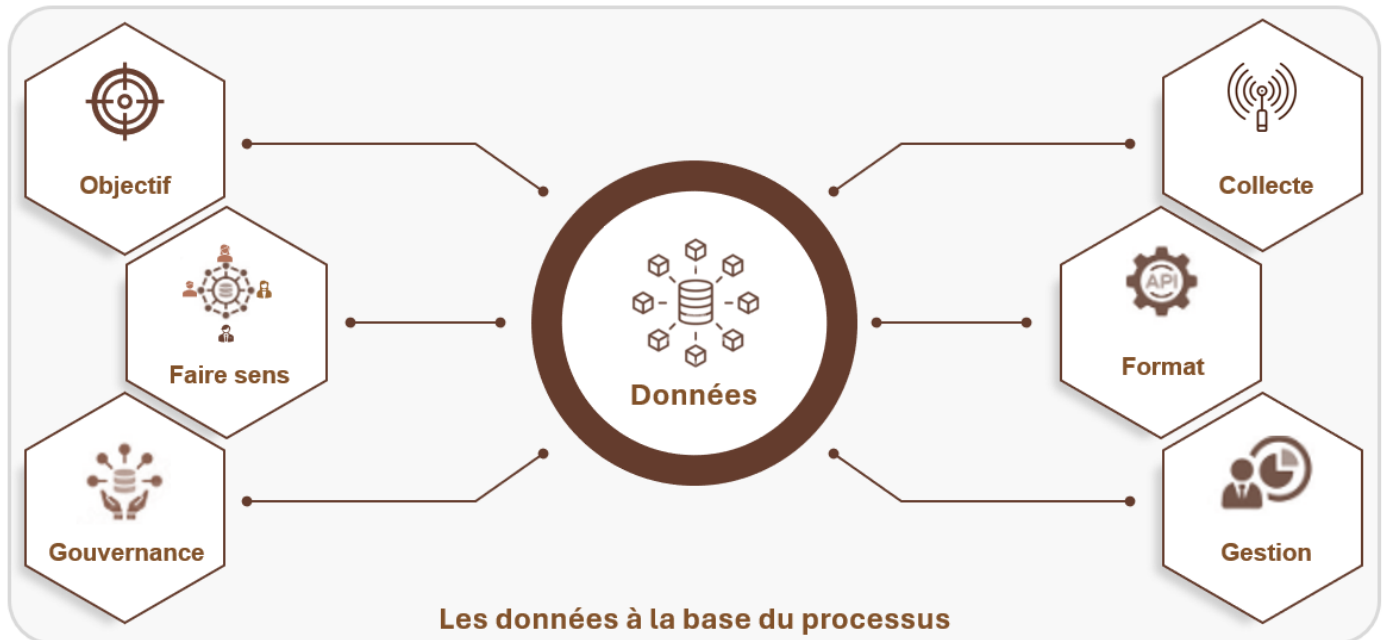
Dès les premières phases d'un projet IA, il faut évaluer la qualité, la structuration et la quantité des données disponibles. Les besoins de retraitement des données sont des paramètres à prendre en compte dans la balance avec les gains attendus pour le cas d'usage.

Bien que les coopératives collectent souvent de grandes masses de données, une partie reste inutilisée faute de besoin identifié. Les données sont au coeur du réacteur qu'elles servent à décrire, à prédire, à prescrire ou à générer. Leur qualité est essentielle, des données incomplètes ou erronées compromettent l'apprentissage des modèles et la fiabilité des résultats.

D'où vient la donnée, à quoi sert-elle ? Chaque métier peut avoir une perception différente de ce qu'est une donnée utile. Une collaboration entre métiers concernés, dès le lancement du projet jusqu'à l'élaboration du cahier des charges pour produire le cas d'usage est donc indispensable pour définir les besoins, les données nécessaires, adaptées, les outils à utiliser et les responsabilités de chacun à chaque étape du développement.

Par exemple, pour suivre la performance de la production laitière en lien avec l'alimentation, à travers un tableau de bord, il faudra identifier les données pertinentes, leur format, le chemin à suivre pour les collecter, les calculs à effectuer et le rôle de l'IA dans la modélisation.

La gouvernance des données permet de clarifier les rôles : qui collecte, qui utilise, qui en est propriétaire...



“ Les données, un enjeu stratégique

« Enfin, la « transition numérique » en agriculture amène des changements devant être traités à l'échelle de politiques publiques. Le point qui nous semble le plus important est la gouvernance de la donnée. Les nouveaux moyens techniques ont transformé les agriculteurs en producteurs de données, que ce soit via des capteurs qu'ils possèdent, ou via des saisies manuelles de leur part (approches participatives). Il y a une réflexion urgente à mener sur la valeur de ces données, leur droit d'usage et leur protection. C'est un sujet difficile car d'une part la mise en commun de la grande masse de données produites peut bénéficier à tous, mais d'autre part trop d'ouverture peut nuire à la compétitivité d'un exploitant ou d'un territoire.

Actuellement, une part de ces données se retrouve entre les mains d'acteurs privés (exemple : les fournisseurs des technologies et des services) qui bénéficient d'une vision globale sans forcément une grande réciprocité. Pour éviter cet écueil, le conseil national du numérique recommande de conférer aux données agricoles un statut de bien commun».

Livre blanc n°6_INRIA et INRAé_Agriculture et numérique_Tirer le meilleur du numérique pour contribuer à la transition vers des agricultures et des systèmes alimentaires durables

Calibrer les données, un premier enjeu

Il pourra être nécessaire de définir un dictionnaire de langage, de données pour la coopérative, définir chaque concept et l'énoncer clairement, trouver un champ lexical avec une définition commune acceptable par tous les métiers. Chacun peut avoir une définition différente des mots clés importants pour les indicateurs de suivi du cas d'usage.

Projet BOVALIM : vers une standardisation des données d'alimentation des bovins laitiers

Le projet BOVALIM a été porté par le CNIEL en partenariat avec GS1 France et AgDataHub dans le cadre de Numagri.

Ce premier cas d'usage vise à standardiser les données d'alimentation du troupeau bovin laitier pour faciliter leur échange. Le projet associe les acteurs concernés, notamment La Coopération Agricole avec ses sections nutrition animale et laitière.

Ce projet s'inscrit dans la stratégie interprofessionnelle de la gestion et de la valorisation des données de la filière laitière française, pour une circulation maîtrisée des données entre les maillons, dans un climat de confiance entre les acteurs et au bénéfice de chacun d'entre eux.

La finalité de ce projet est de construire un langage commun pour améliorer la transparence en augmentant l'efficacité et la fiabilité des échanges pour les acteurs de la nutrition animale et de la filière laitière.

“ La réflexion interprofessionnelle part de plusieurs constats :

- Avec l'essor des outils numériques, de nombreuses bases de données recensent des informations sur le lait, les produits laitiers, les pratiques d'élevage, de fabrication et de commercialisation.
- De plus en plus d'acteurs souhaitent accéder à ces données et proposent des solutions de valorisation.
- De plus en plus d'usages sont recensés pour une même donnée.

CNIEL

Standardiser les données est un prérequis pour faciliter leur collecte, leur circulation et leur valorisation. Il s'agit de disposer de dictionnaires communs pour la collecte des données et de standards d'échange harmonisés pour leur

circulation (par exemple un code-barres). Cette standardisation favorise l'interopérabilité des outils numériques.

Les bénéfices attendus sont :

1. Gagner en efficacité à tous les maillons de la filière

Simplifier le suivi de démarches, réduire la charge administrative, faciliter le suivi des fournisseurs et des clients

Optimiser la collecte, les contrôles et le partage des informations liées à l'alimentation

Alimenter des outils prospectifs et de suivis

2. Valoriser les pratiques

3. Faciliter la traçabilité, en simplifiant et fiabilisant le passage d'informations tout au long de la chaîne jusqu'au consommateur, avec une meilleure mise en avant des bonnes pratiques et des démarches de segmentation (non déforestant, sans OGM, AOP, bio, etc.).

Le préalable pour la bonne valorisation est que les données répondent aux principes directeurs des « **données FAIR** » (**Findable, Accessible, Interoperable et Reusable**) qui garantissent la capacité des systèmes informatiques à trouver, interopérer et réutiliser des données avec une intervention humaine minimale.

Usage des données

Avant de déployer une solution technique, il est essentiel de définir précisément les données à collecter (leur nature, leur fréquence, leur objectif, le niveau de précision attendu...). Ces éléments qui émanent de la définition des besoins varient fortement selon les usages et orientent le choix des solutions techniques à tout niveau (capteurs, matériels, logiciels, interfaces, etc.). Il faut ensuite s'interroger sur la manière de capter cette donnée. Enfin, l'échantillonnage dépendra de l'utilisation finale de la donnée, notamment si elle sert à modéliser ou entraîner une IA.

L'exemple de l'outil Farm-e-val développé par CCPA illustre cette démarche. Les données à collecter au service de l'optimisation, des performances et du bien-être ont été sélectionnées pour répondre à 3 objectifs :

1. Piloter l'innovation en nutrition,
2. Accompagner le conseil aux éleveurs,
3. Comparer pour un conseil stratégique.

Selon les cas, il pourra s'avérer nécessaire de s'assurer que les données pourront répondre à l'usage de modélisation ou d'entraînement de l'IA.

Par exemple si les données sont issues de capteurs, le parcours de choix et de validation des capteurs pourra être le suivant :

- Veille technologique, sélection du capteur,
- Tests et validation du capteur hors site et sur site pour validation des mesures et possibilité de traitement selon utilisation attendue, validation de la fiabilité de la remontée des données en situation réelle, validation de l'utilité par rapport aux objectifs fixés,
- Tests terrain avec utilisateurs pour valider l'usage, optimiser les réglages, le positionnement, la durée de vie des capteurs in situ,
- Intégration des données pour modélisation ou entraînement.

Outre les biais de mesure relatifs aux outils de collecte des données, il est également crucial d'anticiper les biais pouvant affecter la qualité des données :

- **Biais de sélection** : s'assurer de la représentativité des données, population suffisamment large,
- **Biais de confirmation**: s'assurer que l'algorithme n'a pas été fait de manière à arriver au résultat auquel le développeur s'attend.
- **Biais d'observation** : s'assurer que la méthode de collecte n'influence pas les données,
- **Biais de rapportage** : s'assurer que les observations ne sont pas rapportées de manière disproportionnée,
- **Biais d'exclusion** : s'assurer que certaines caractéristiques ne sont pas sous-estimées.

Collecter les données

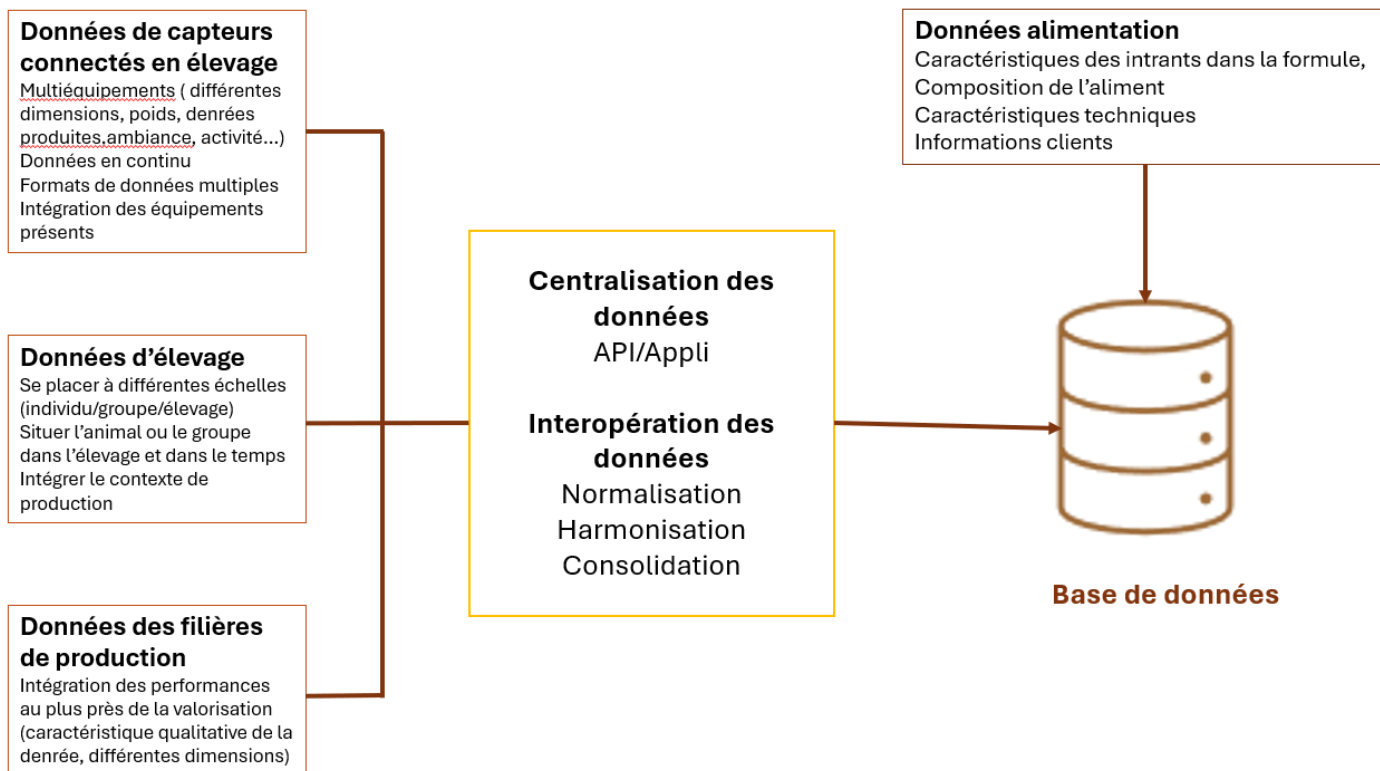
La maîtrise de l'accès aux données et leur valorisation représentent un atout majeur pour l'ensemble des acteurs de la filière agricole, qu'il s'agisse d'enjeux partagés ou spécifiques. Le modèle coopératif notamment lorsque les associés coopérateurs produisent à la fois du végétal et des produits animaux, tout en disposant d'outils de transformation comme les usines d'alimentation animale, met les adhérents dans une position privilégiée pour la collecte de données.

“ « Le partage des données agricoles est une priorité à la fois pour la profession agricole et pour la recherche agronomique afin de soutenir le développement des connaissances agronomiques et des technologies et services numériques en agriculture. C'est également un enjeu fort de souveraineté numérique. »

Livre blanc n°6_INRIA et
INRAé_Agriculture et numérique_Tirer le meilleur du numérique pour contribuer à la transition vers des agricultures et des systèmes alimentaires durables

Les données collectées peuvent provenir de multiples sources et doivent être transmises efficacement. Si certains systèmes utilisent des liaisons filaires, d'autres utilisent des capteurs dotés de technologies sans fil. La captation et la transmission de la donnée en agriculture s'appuie de plus en plus sur l'Internet des objets (IoT), en particulier la RFID et les réseaux de capteurs sans fil, avec des spécificités propres au secteur.

Le niveau de précision ou la fréquence des mesures influence directement le volume de données à transmettre, impliquant un arbitrage entre performance, efficacité, précision et coût.



Notion de consentement éclairé

Pour instaurer un climat de confiance entre les parties, il est essentiel de définir en amont les conditions d'utilisation des données échangées.

Concernant les données personnelles ou mixtes, les coopératives sont soumises au Règlement Général sur la Protection des Données (RGPD), en vigueur depuis le 25 mai 2018¹. Elles appliquent déjà cette réglementation dans le cadre des traitements de données réalisés, notamment vis-à-vis des coopérateurs.

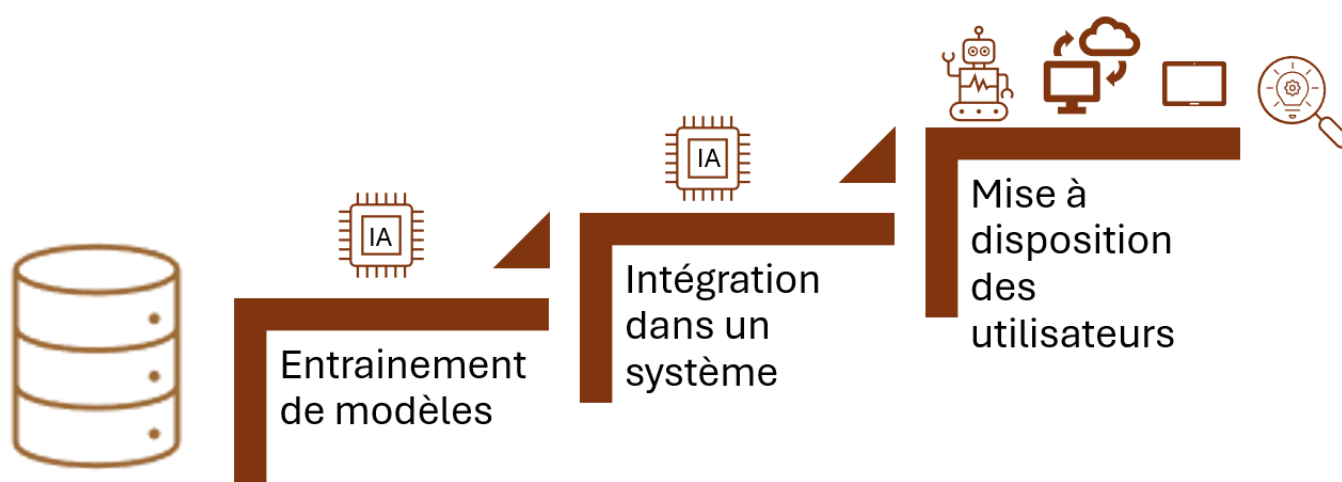
Le consentement de la personne concernée par le traitement de la donnée est obligatoire. Il doit être **libre, spécifique, éclairé et explicite**. En cas de finalités multiples, un consentement distinct doit être recueilli pour chacune d'elles, après une information claire et compréhensible. Un consentement global n'est pas valable si les finalités ne sont pas précisément définies. De plus, toute évolution des finalités en cours de traitement nécessite un nouveau consentement.

Un consentement comporte plusieurs informations :

- L'acteur qui donne son consentement,
- Un bénéficiaire du consentement,
- Une famille de données concernées,
- Un usage,
- Une période avec une date de début et une date de fin.

Transformer les données en informations

Selon le cas d'usage choisi, l'atteinte de l'objectif repose sur l'alliance entre données connectées et intelligence des données, mobilisant des outils de biostatistiques, de modélisation, d'intelligence artificielle comme le Machine Learning (voir exemple des fermes connectées). Ces technologies permettent d'explorer les dynamiques d'évolution, de comprendre les interactions entre performance, nutrition et environnement, de prédire, détecter précocement, de construire des références techniques, des historiques et des comparatifs.



L'IA ouvre ainsi des perspectives inédites : suivi dynamique, analyses individualisées, gestion de la complexité des données. Toutefois, elle ne résout que les problèmes pour lesquels elle a été conçue.

La réussite dépend de plusieurs facteurs : quantité et qualité des données, puissance et temps de calcul nécessaires, et expertise humaine pour valider chaque étape. Les choix technologiques et méthodologiques doivent être adaptés à chaque coopérative, ses besoins et ses atouts.