

LE GUIDE TRAKMY · DONNÉE & IA

Préparer ses données d'actifs pour l'IA

Comment passer d'un parc connecté à une donnée réellement exploitable : référentiel, six dimensions de qualité, gouvernance, architecture et feuille de route en 90 jours pour faire de votre parc une source de données fiable.

85 %

DES PROJETS IA ÉCHOUENT (DONNÉE)

6

DIMENSIONS DE QUALITÉ

90 jours

FEUILLE DE ROUTE

01 La donnée, pas le modèle

Pourquoi 85 % des projets IA échouent

02 Une donnée « prête pour l'IA »

5 propriétés, 3 états de maturité

03 Le référentiel d'actifs

La fondation que tout le monde néglige

04 Les 6 dimensions de qualité

Mesurer, suivre, corriger

05 La télémétrie

Séries temporelles, position, capteurs

06 Gouvernance, RGPD, souveraineté

Qui possède, qui accède, où héberger

07 L'architecture de la donnée

De la collecte à la décision

08 Les cas d'usage IA

Surparc, anomalies, maintenance

09 La feuille de route 90 jours

Cadrer, fiabiliser, exploiter

10 Pièges & checklist

Vérifier que le parc est prêt

● L'ESSENTIEL

Le modèle n'est presque jamais le maillon faible

RÉPONSE DIRECTE

Une donnée d'actif « prête pour l'IA » n'est pas une donnée abondante : c'est une donnée alignée sur un cas d'usage, identifiée de façon unique, gouvernée, alimentée par des flux contrôlés et vérifiée en continu.

85 %

Des échecs IA = donnée (Gartner)

> 80 %

Des projets IA jamais en prod (RAND)

42 %

Ont abandonné l'IA en 2025 (S&P)

Le vrai problème

Équiper 300 engins de traceurs crée un gisement de données — mais tant qu'elle n'est pas structurée, identifiée et gouvernée, elle reste « un point sur une carte ». Le problème n'est presque jamais l'absence de données, mais leur **état**.

Un cas particulier

La donnée d'un parc est **temporelle** (l'instant et l'ordre comptent), **spatiale** (rapportée à des zones) et **massive et continue**. Trois propriétés que la gestion de données classique ne couvre pas.

Le coût caché

Plus d'un quart des organisations perdent plus de 5 M\$/an à cause d'une donnée de mauvaise qualité (IBM, 2025) : décisions sur un taux d'utilisation faux, surparc maintenu, alertes ignorées, maintenance qui ne prédit rien.

« Investir dans la donnée avant l'algorithme inverse le ratio d'échec. »

Sources : Gartner 2025 ; RAND Corporation ; S&P Global 2025 ; IBM Institute for Business Value 2025. Chiffres externes datés, indicatifs.

Ce qu'est une donnée « prête pour l'IA »

RÉPONSE DIRECTE

Une donnée AI-ready est *fit-for-purpose* : ni la plus volumineuse, ni la plus récente — celle dont un modèle ou un tableau de bord peut se servir sans retraitement manuel.

Les cinq propriétés à vérifier une à une

- ✓ **Alignée sur un cas d'usage** : vous savez à quelle question elle répond (surparc, panne, tournée, preuve).
- ✓ **Gouvernée au niveau de l'actif** : chaque actif a une identité unique, un propriétaire, des règles d'accès.
- ✓ **Alimentée par des pipelines à points de contrôle** : validité contrôlée à chaque étape.
- ✓ **Décrite par des métadonnées vivantes** : origine, signification, unité, date — sinon une valeur n'est qu'un nombre orphelin.
- ✓ **Assurée en continu** : la qualité est surveillée en permanence, pas vérifiée une seule fois.

Trois états de maturité

Donnée brute

La trame qui sort du traceur : ID technique, coordonnées, horodatage, capteurs. **Exacte mais muette.**

Donnée structurée

Rapprochée du référentiel : actif nommé + catégorie, présence en zone, temps normalisé. **Exploitable.**

Donnée enrichie

Croisée avec d'autres sources (comptable, planning, maintenance). **Prête pour l'IA** — c'est là qu'est la valeur.

TEST SIMPLE

« Si je donnais ce fichier à un analyste qui ne connaît pas mon parc, pourrait-il l'exploiter sans me poser de questions ? » Si non, la donnée n'est pas prête — un modèle est encore moins indulgent.

Le référentiel d'actifs

RÉPONSE DIRECTE

Si vous ne devez faire qu'une chose avant de parler d'IA : bâtir et tenir un référentiel propre. La quasi-totalité des projets butent sur l'absence de ce socle — la donnée existe, mais n'est rattachée à rien.

Un identifiant pivot, neutre et immuable

Chaque actif possède un identifiant unique qui ne change **jamais**, même s'il est réaffecté ou ré-équipé. On y rattache tous les autres identifiants (n° de série, code comptable, immatriculation) par une table de correspondance. Le traceur change ? On met à jour la correspondance, l'historique reste intact.

Une taxonomie fermée

Catégorie, sous-catégorie, modèle : choisis dans une **liste fermée**, jamais en texte libre. « Pelle / pelleteuse / mini-pelle » désigne le même engin sans qu'aucun traitement ne puisse les regrouper.

Attribut	Pourquoi il débloque l'analyse
Identifiant pivot	Réconcilie les sources, garantit l'historique
Catégorie (liste fermée)	Permet l'agrégation entre familles
Valeur d'acquisition	Calcule le ROI, pondère les alertes
Date de mise en service	Calcule l'âge, base de la maintenance
Affectation / site	Détecte les écarts théorie / terrain
Statut	Distingue immobilisation normale et anomalie

SIGNAL D'ALERTE

Un référentiel qui vit dans un tableur partagé, mis à jour à la main sans règle, est non fiable par défaut. Un référentiel sérieux a un **propriétaire désigné**, des règles de saisie, un historique et une source de vérité unique.

Les six dimensions de la qualité

RÉPONSE DIRECTE

« Qualité » est trop vague pour être actionnable. On la décompose en six dimensions mesurables. Une seule défaillante suffit à fausser un modèle – l'IA propage les erreurs au lieu de les corriger.

Dimension	Ce qu'elle mesure	Comment la contrôler
Complétude	Part des données attendues réellement présentes	% par actif et par période ; seuil de suspicion
Exactitude	Écart entre la donnée et la réalité	Règles de plausibilité, recoupement, échantillon terrain
Cohérence	Absence de contradictions intra/inter-sources	Croisements ; sémantique et unités stables
Fraîcheur	Délai production → disponibilité	Adapter la cadence au besoin (≠ la plus fraîche possible)
Unicité	Aucun doublon d'actif, d'événement, de mesure	Identifiant pivot + déduplication ID + horodatage
Traçabilité	On sait d'où vient la donnée et ses transformations	Métadonnées de lignage tout au long du pipeline

À RETENIR

Les dimensions faibles ne se compensent pas : un modèle bute sur la plus mauvaise. Construisez le **profil de qualité** de votre parc et corrigez en priorité ce qui passe sous le seuil (souvent la cohérence et la traçabilité).

Séries temporelles, position et capteurs

RÉPONSE DIRECTE

La télémétrie décrit le comportement réel de l'actif. C'est la donnée la plus délicate : continue, irrégulière et sensible aux conditions terrain.

- ✓ **Normaliser le temps** : tout en référence unique (UTC stocké, affichage local). Une série au temps faux est inexploitable.
- ✓ **Gérer les trous** : stockage hors-ligne pour combler en différé ; **marquer** les périodes sans signal, ne jamais interpoler en silence.
- ✓ **Position ≠ présence en zone** : conserver la nature (GPS, Wi-Fi, BLE, cellulaire) et la précision de chaque point.

Grandeur	Cadence pertinente
Position (actif passif)	1 à 4 / jour + événement mouvement
Position (temps réel)	Continue en roulage
Activité / vibration	Échantillonnée (seuil + agrégation)
Choc	Événementielle (seuil)
Température / humidité	1 / heure + alerte seuil
Ouverture / contact	Événementielle

LE BON RÉFLEXE

Une donnée interpolée non signalée est **pire qu'un trou** : le trou dit « je ne sais pas », l'interpolation masquée affirme une fausseté. La cadence la plus utile est celle alignée sur la dynamique du phénomène, pas la plus fréquente (qui vide la batterie).

RGPD, conservation et souveraineté

RÉPONSE DIRECTE

Une donnée gouvernée est une donnée qu'on peut exploiter en confiance, à l'échelle et dans la durée. La gouvernance n'est pas une couche ajoutée après coup : c'est une propriété de la donnée AI-ready.

- ✓ **Propriété** : chaque jeu de données a un propriétaire métier (qualité, règles d'usage), distinct du responsable technique.
- ✓ **RGPD** : dès qu'un actif suit indirectement une personne (véhicule conduit par un salarié), finalité explicite, information, minimisation, durées et droits s'appliquent (cadre CNIL).
- ✓ **Conservation** : ne garder que le nécessaire, le temps nécessaire – durées paramétrables par type de donnée.

Souveraineté = sécurité économique

La concentration de toutes les positions d'une flotte dessine la **carte opérationnelle** de votre entreprise. Un hébergement en **Union européenne** protège des législations extraterritoriales. Son corollaire : la **réversibilité** – pouvoir récupérer l'intégralité de vos données et quitter un fournisseur sans blocage.

SIGNAL D'ALERTE

Une donnée personnelle collectée hors cadre (finalité non tracée, personnes non informées) est **inexploitable** : le jour du projet, vous n'aurez pas le droit de l'utiliser. La conformité se prépare en amont.

De la collecte à la décision

RÉPONSE DIRECTE

Comprendre la chaîne de bout en bout permet de localiser où la qualité se gagne et se perd. La valeur se joue surtout aux premiers maillons : une donnée mal collectée n'est jamais rattrapée par un beau tableau de bord.

- 1 Collecte**
Le traceur capte position et mesures. Qualité = matériel, configuration (cadence, seuils), robustesse réelle.
- 2 Transport**
Le réseau achemine (NB-IoT/LTE-M pour l'autonome, 4G pour le temps réel). Zones blanches et stockage hors-ligne se jouent ici.
- 3 Ingestion & stockage**
Réception, déduplication, horodatage : c'est là que se construisent unicité et traçabilité.
- 4 Transformation & qualité**
Nettoyage, normalisation, enrichissement par le référentiel, contrôle des 6 dimensions. L'étape qui produit la donnée AI-ready.
- 5 Exposition**
Tableaux de bord, exports et surtout interfaces machine (API, MQTT) pour les outils analytiques et les modèles.

L'API : la porte d'entrée

Une **API ouverte, documentée et bidirectionnelle** (format JSON) est le critère technique le plus déterminant : sans elle, la donnée reste prisonnière de la plateforme et chaque analyse devient un export-import manuel.

MQTT & système d'information

Le **MQTT** (et les webhooks) pousse les événements en temps quasi réel pour les cas réactifs. La valeur naît du croisement avec l'ERP, le TMS, le SIG, la GMAO et l'export **Power BI**.

Ce que des données propres débloquent

RÉPONSE DIRECTE

Inutile de viser des modèles exotiques tant que les cas d'usage fondamentaux ne sont pas servis. Quatre familles créent le plus de valeur – à aborder dans l'ordre de la maturité de données requise.

Cas d'usage	Gain typique (études 2025-26)	Dimensions critiques
1. Optimisation / surparc	-10 à -15 % de surparc	Complétude, cohérence (référentiel)
2. Détection d'anomalies / vol	Récupération accélérée, moins de pertes	Cohérence, fraîcheur, unicité
3. Maintenance prédictive	-30 à -50 % d'immobilisations, ROI ~10:1	Exactitude, complétude, traçabilité
4. Pilotage prédictif	Décisions d'investissement fiabilisées	Toutes

L'ORDRE COMPTE

Commencez par le **surparc** (gain rapide, peu exigeant), puis les **anomalies** (peu d'historique requis). Ces deux étapes financent la suite et accumulent l'historique propre dont la **maintenance prédictive** a besoin. Commencer par le cas le plus exigeant est la 1^{re} cause d'échec.

Sources : US Department of Energy ; études maintenance prédictive 2025-2026 ; McKinsey. Chiffres indicatifs, à valider sur votre périmètre.

90 jours vers une donnée prête pour l'IA

RÉPONSE DIRECTE

La pire stratégie : tout nettoyer avant de rien exploiter. La bonne : itérative et orientée usage – choisir un cas d'usage, préparer juste ce qu'il faut, livrer, mesurer, étendre.

- 1 Jours 1-30 · Cadrer & fonder**
Choisir un cas d'usage prioritaire et chiffrer l'enjeu ; construire/nettoyer le référentiel ; cartographier les sources ; établir le profil de qualité initial (6 dimensions).
- 2 Jours 31-60 · Fiabiliser & mesurer**
Corriger les dimensions sous le seuil ; mettre en place les contrôles automatiques (plausibilité, déduplication, marquage des trous, temps) ; documenter les métadonnées ; suivi qualité continu.
- 3 Jours 61-90 · Exposer & exploiter**
Exposer par l'API ; enrichir avec une source métier ; lancer le pilote analytique ; mesurer contre le critère de succès, puis décider de l'extension.

CONSEIL

Les organisations qui réussissent leurs projets d'IA ont deux fois plus de chances d'avoir **repensé leur processus métier avant de choisir leurs outils** (McKinsey, 2025). Commencez par la question métier et la donnée, pas par la technologie.

Vérifier que votre parc est prêt

RÉPONSE DIRECTE

Cinq pièges récurrents ruinent les projets de données d'actifs. La checklist permet de vérifier, point par point, qu'un parc est réellement prêt pour l'IA.

Les 5 pièges

- ✓ **Collecter tout, exploiter rien** (pas de cas d'usage prioritaire).
- ✓ **Sous-estimer le référentiel** (1^{re} cause d'échec).
- ✓ **Masquer l'incertitude** (interpoler les trous en silence).
- ✓ **Reporter la conformité** (donnée personnelle hors cadre = inexploitable).
- ✓ **Accepter l'enfermement propriétaire** (ni export, ni reprise).

Checklist – mon parc est-il prêt ?

- ✓ Référentiel : identifiant pivot, taxonomie fermée, attributs clés, propriétaire désigné
- ✓ Qualité : 6 dimensions suivies, trous marqués, horodatages normalisés, déduplication
- ✓ Gouvernance : finalité documentée, personnes informées (RGPD), conservation définie, hébergement UE + réversibilité
- ✓ Architecture : API ouverte, flux temps réel (MQTT), ponts SI (ERP/TMS/GMAO/Power BI), lignage

Comment TRAKmy fournit une donnée prête pour l'IA

RÉPONSE DIRECTE

TRAKmy est conçu pour produire une donnée exploitable, pas seulement des points sur une carte. Chaque maillon du guide trouve sa réponse native.

- ✓ **À la source** : matériel robuste (IP68/69K, IK09, jusqu'à 10 ans) conçu et fabriqué en Europe. ULTRA N+ (1101) et ses 5 capteurs produisent les séries d'usage de la maintenance prédictive ; COMPACT (1110) pour l'outillage ; REACT (1203) pour le temps réel et le kilométrage.
- ✓ **Au transport** : NB-IoT (70+ pays), stockage hors-ligne qui transforme un trou définitif en trou comblé en différé.
- ✓ **Au stockage & à la qualité** : plateforme (Essential / Professional / Analytics), géo-zones, alertes, conservation paramétrable (12 / 24 / 36 mois). Ancres Geobeacon pour la continuité indoor-outdoor.
- ✓ **À l'exposition** : API, MQTT (dès Professional), export Power BI (Analytics).
- ✓ **À la gouvernance** : données hébergées en Europe (UE), support français, réversibilité garantie.

EN RÉSUMÉ

TRAKmy fournit la **chaîne fiable de bout en bout** ; la méthode de ce guide vous donne le référentiel, les six dimensions et la feuille de route. Pour une réponse chiffrée sur votre parc, demandez un devis.

● ANNEXE · À EMPORTER

Les questions à poser à votre fournisseur

RÉPONSE DIRECTE

À poser dès l'évaluation : elles séparent les plateformes conçues pour être **intégrées** de celles conçues pour être seulement regardées.

Collecte & matériel

- ✓ Quelle autonomie réelle pour *ma* cadence (pas le catalogue) ?
- ✓ Quels capteurs, à quelle cadence configurable ?
- ✓ Quel taux de défaillance terrain documenté sur plusieurs années ?

Transport & intégrité

- ✓ Combien de temps le traceur stocke-t-il hors-ligne, remontée garantie sans perte ?
- ✓ Comment les périodes sans signal sont-elles marquées ?
- ✓ Les doublons de trames sont-ils dédupliqués automatiquement ?

Exposition & intégration

- ✓ API ouverte, documentée, bidirectionnelle ? Sous quel format ?
- ✓ MQTT / webhooks pour le temps réel ?
- ✓ Export Power BI ? Durée de conservation paramétrable ? Récupération totale à la sortie ?

Gouvernance

- ✓ Où mes données sont-elles physiquement hébergées ?
- ✓ Comment respecter le RGPD si je suis des véhicules conduits par des salariés ?

PASSEZ À L'ACTION

Faisons de votre parc une source de données fiable

Matériel fiable à la source, stockage hors-ligne, conservation paramétrable, API, MQTT et export Power BI, données hébergées en Europe : TRAKmy fournit la chaîne AI-ready de bout en bout. Demandez une démonstration et un devis.

[Demander un devis](#)